



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 141465

TO: Jon E Angell

Location: REM-2C18 2d20

Art Unit: 1635

Thursday, January 06, 2005

Case Serial Number: 09/701618

From: Noble Jarrell

Location: Biotech-Chem Library

Rem 1B71

Phone: 272-2556

Noble.jarrell@uspto.gov

### Search Notes

**This Page Blank (uspto)**

141465

STIC-Biotech/ChemLib

From: Angell, Jon E  
Sent: Wednesday, December 29, 2004 11:14 AM  
To: STIC-Biotech/ChemLib  
Subject: RE: Sequence Database Search Request

My apologies for the typo...

The correct serial number is: 09/701,618 CRFE ✓

Thanks,  
Eric

-----Original Message-----

From: STIC-Biotech/ChemLib  
Sent: Wednesday, December 29, 2004 11:11 AM  
To: Angell, Jon E  
Subject: RE: Sequence Database Search Request

There is no valid CRF for this serial number, please provide us with another valid serial number.(in our in-house database)

Thank you.  
Gary

-----Original Message-----

From: Angell, Jon E  
Sent: Tuesday, December 28, 2004 4:01 PM  
To: STIC-Biotech/ChemLib  
Subject: Sequence Database Search Request

SEARCH REQUEST FORM  
Scientific and Technical Information Center

Examiner# : 78697  
Art Unit : 1635  
Phone Number: 571-272-0756  
Date: 12/28/04  
Serial Number: 09/701,608 Rothbarth et al.  
Mailbox & Bldg/Room Location: REMSEN 2C18  
Results Format Preferred (circle): Paper

\*\*\*\*\*

STAFF USE ONLY

Searcher: Noble  
Searcher Phone: 2-  
Date Searcher Picked up:  
Date Completed: 1/16/04  
Searcher Prep/Rev. Time: to  
Online Time: to

\*\*\*\*\*

Type of Search

NA Sequence: # 4  
AA Sequence: # 2  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: COMPUBEN  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**This Page Blank (uspto)**



I would like to have a standard search performed using the following SEQ. ID NOs. from application : 09/701,608

SEQ ID NO. 1 (DNA ~1156 nucleotides long)

SEQ ID NO. 2 (Polypeptide ~141 amino acids long)

SEQ ID NO. 3 (DNA ~1040 nucleotides long)

SEQ ID NO. 4 (Polypeptide ~141 amino acids long)

Please perform a standard nucleic acid search for SEQ ID NO. 1 AND SEQ ID NO. 3

Please perform a standard polypeptide search for SEQ ID NO. 2 AND SEQ ID NO. 4

AND also perform a search for any nucleic acid sequences which encode SEQ ID NO.2 and SEQ ID NO.4

NOTE: SEQ ID NO. 1 is identified as encoding SEQ ID NO. 2

SEQ ID NO. 3 is identified as encoding SEQ ID NO. 4

and SEQ ID NO. 2 is highly similar to SEQ ID NO. 4 (differs by 14 amino acids) and they have the same function

Thanks,  
Eric

*J. Eric Angell*

Art Unit 1635

Office: REMSEN 2D20

mailbox: REM 2C18

571-272-0756

\*\*\*\*\*

STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2- \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA Sequence: # \_\_\_\_\_  
AA Sequence: # \_\_\_\_\_  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**This Page Blank (uspto)**



## FEATURES

**Source**

```
Location/Qualifiers
1. 1156
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
```

## ORIGIN

### Alignment Scores:

Pred. No.:	6	77e-69	Length
Score:	715.00		Match
Percent Similarity:	100.00%		Cons
Best Local Similarity:	100.00%		Mism
Query Match:	100.00%		Indel
DB:	6		Cons

US-09-701-618A-2 (1-141) X BD221634 (1-1156)

QY	118	ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTGTTCAGCG	177
QY	21	PheGluSnsSerIleGlyAlaValAAspGluMetLeuIysThrMetMetSerValSerArg	40
Db	178	TTTGAGAAATTCATGGTGTGCTGTGGATGAGATGCTGAAGACCATGATATCTCGTTTCAGA	237
QY	41	AsnGluLeuLeuGlnIlyLeuAspProLeuGlnAlaIlyValAspLeuValSerAla	60
Db	238	AATAGTGTGTGACAGAGTTGATTCACATTGAACAGCAAAAGTGATTTGGTTCTTCGA	297
QY	61	TyThrLeuSnsSerMetPheTrpValIlyTrpLeuAlaTrpGlnGlyValAspProIySglu	80
Db	298	TACACATTAATTAATTCATGTTGTGGTTATTTGGCAACCCAGGAGTTAATCCTAAGAA	357
QY	81	HisProValIySglnGluLeuGlnIlyLeuArgIlyLeuValIlyMetAsnArgValIlySglnIle	100
Db	358	CATCCAGTAAACAGGAAATTTGGAAGAAACAGAGATATTAACACAGATCAAGGAAAAA	417
QY	101	ThrAspIlyValIlySalaGlyIlyLeuSnsAspArgGlyAlaIleAsnSerArgPheValIlySns	120
Db	418	ACGACACAAAGAAAAAGCTGGCAAGCTGGACAGAGGTGCAGCTTCMAAATTTGTAAAAAAT	477
QY	121	AlaLeuTrpGluProIySnsIlyAsnAlaSerIlyValAlaAsnIlyGlyIlySnsIyS	140
Db	478	GCCTCTGGGAACCAAAATCGAAAAATCGAAATATGCAAAAGTTGGCAATAAAGGAAAAAATAAA	537
QY	141	Ser 141	
Db	538	AGT 540	

RESULT 2		
AX009361		
LOCUS		
DEFINITION	AX009361	1156 bp
DESCRIPTION	Sequence 1 from Patent WO963071.	DNA linear
		PAT 06-SEP-2000

## REFERENCE

**AUTHORS**  
**TITLE**  
**JOURNAL**

DEUTSCHES KREBSFORSCH (DE) ; ROTHBARTH KARSTEN (DE) ; WERNER DIETER  
(DE) ; STAMMER HERMANN (DE)

## FEATURES

## Source

```
source
1. location/Qualifiers
1. .1156
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
118. .543
/note="unnamed protein product"
```

CDS

## ORIGIN

**Alignment Scores:**

Pred. No.:	6, 77e-69	Length
Score:	715.00	Match
Percent Similarity:	100.00%	Cons
Best Local Similarity:	100.00%	Mism
Query Match:	100.00%	Indel
GB:	6	Gaps

US-09-701-618A-2 (1-141) X AX009361 (1-1156)

QY	1	MetAlaGlyGluGluIuileAenGluAspTyrProValGluIleHsgIuTyrLeuSerAla	20
Db	118	ATGGCAGGTGAAAGAAATTAAATGAAACATCTCCAGTAAATAATTCACAGTATTGTCAAGC	177
QY	21	PheGluAsnSerIleGlyValAvalAspGluMetLeuSerThrMetMetSerValSerArg	40
Db	178	TTTGAGAAATTCATTGGTGGCTGTGTGATAGATGCTGAAGACCAAGAAAGTCTGTTTCAGA	237
QY	41	AenGluLeuLeuGluLysLeuAspProLeuGluIuAluysValAspLeuValSerAla	60
Db	238	AATGAGTTGTGCAGAAAGTTGATGCACCTTGAAACAAAGAGTATTGGTTTTCGCA	297
QY	61	TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu	80
Db	298	TTCAATTAATTTCAATCTTTTGGGTTTAAATTTGGCAACCCAAAGAGTTAATCTTAAGGAA	357
QY	81	HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle	100
Db	358	CATCAGATTAACACAGAAATTCGAAAGAAATCAGATATATTTGAACAGAGCTCAACAGGAAATTA	417
QY	101	ThrAspLysLysLysAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn	120
Db	418	ACAGACAAAGAAAAAGGCTGGCAGCTGGACAGAGGTGCAGGCTTCAAGATTGTGAAAAAAT	477
QY	121	AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys	140
Db	478	GCCTCTGGAAACCAAAATCGAAAAATGCATCAAAAAGTTCCAATTAAGAAAAAAGTTAA	537
QY	141	Ser 141	
Db	538	AGT 540	

RESULT 3	LOCUS	DEFINITION	ACCSSION
BC016284	1168 bp	mRNA	PRI 29-JUN-2004
		Homo sapiens nuclear DNA-binding protein, transcript variant 1,	
		mRNA (cDNA clone MGC:9308 IMAGE:3306517), complete cds.	
		BC016284	

## REFERENCE

## AUTHORS

```

/codon_start=1
/protein_id="CAC07450.1"
/db_xref="GI:1996663"
/translation="MGGESEINDEYPIAEIHEYSAAEENISGADEMLKTWMSVSNELL
OKLDEQAYDLSAYATLLSMSEWVLAQVNPKEHPEKQELERIRVYNNRVEIETD
KKKKGKILDRGAASRPVNNLWSEKSSKNSKVNKSKSSK"

```

Klausner, R. R., Collins, F. S., Wagner, L., Stemann, C. M., Schuler, G. I., Altschul, S. P., Zeeberg, B., Buecaw, K. H., Schaefer, C. F., Bhat, N. K., Hopkins, R. P., Jordan, H., Moore, I., Max, S. I., Wang, J., Hsieh, F., Datsenko, I., Matsumura, K., Farmer, A. A., Rubin, G. M., Hong, L., Stapleton, M., Soares, M. B., Donald, M. F., Casavante, T. L., Scheetz, T. E., Brownstein, M. J., Usdin, T. B., Toshiyuki, S., Carninci, P., Prange, C. K., Kana, S. S., Loquellano, N. A., Peters, G. J., Adamson, R. D., Muliyil, S. J., Bosak, S. A., McEwan, P. J., Adamson, K. J., Malek, J. A., Gunaratne, P. H., Richards, S.

Pred. No.:	6.85e-69	Length:	1168
Score:	715.00	Matches:	141
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-701-618A-2 (1-141) x BC016284 (1-1168)

Qy	1	MetAlaGlyGluGluIleAengluAapTyTProValGluIleHisGluTyTLeuSerAla	20
Db	49	ATGGCAGTAGAAGAAATTAATGAAGACTATCCAGTAGAATAATTCACGAGTATTTGTCA	108
Qy	21	PhoGluAsnSerIleGlyAlaValAAspGluMetLeuTyThrMetMetSerValSerArg	40
Db	109	TTTGAGAAATTCATGGTGCTGGAGATGAATGCTGAAACCATGATGCTGTTCTTCA	168
Qy	41	AsnGluLeuLeuGluIleuLeuAapProLeuGluGluIleAlaIleValAAspLeuValSerAla	60
Db	169	AATGAGTGTTCGACGAAGTTGCATCCATTCGACACAGCAAAAGTGATTTGGTTCTGCA	228
Qy	61	TyTThrLeuAsnSerMetPheTTPValTyTLeuAlaThrGlnGlyValAAsnProLysGlu	80
Db	229	TACACATTAATTAATCAATGTTTGGGTTATTTGGCAACCCAGAGATTAACTCTTAAGAA	288
Qy	81	HisProValLysGlnGluLeuGluAlaArgIleArgValTyTMetAsnArgValLysGluIle	100
Db	289	CATCCAGTAAACAGCAATGGAAAGATCATGATATATGAAACAGATCCACAGGAATA	348
Qy	101	ThrAspLysLysLysAlaGlyLysLeuAspArgGlyValAlaIleSerArgPheValLysAsn	120
Db	349	ACGACACAGAAAGAGGCTGGACAGCTGGACAGAGTGCCACCTTCAGATTTGTAAGAAAT	408
Qy	121	AlaLeuTTPGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys	140
Db	409	GCCTCTGGGACCAAAATGCAAAAATGCATCAAAAGTTCCATATAAGGAAAAAGTAAA	468
Qy	141	Ser 141	
Db	469	AGT 471	

RESULT 4

AX329809	1172 bp	DNA	linear	PAT 09-JAN-2002
LOCUS	AX329809			
DEFINITION	Sequence 318 from Patent WO0194629.			
ACCESSION	AX329809			
VERSION	AX329809.1	GI:18102787		
KEYWORDS				
SOURCE				
ORGANISM	Homo sapiens (human)			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE				
AUTHORS	Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,			
	Horrigan, S., Sopper, D.R. and Weaver, Z.			
	Cancer gene determination and therapeutic screening using signature			
	gene sets			
JOURNAL	Patent: WO 0194629-A 318 13-DEC-2001;			
	Avalon Pharmaceuticals (US)			
FEATURES				
Source	Location/Qualifiers			
	1..1172			
	/organism="Homo sapiens"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:9606"			

ORIGIN

Alignment Scores:

Pred. No.:	6.88e-69	Length:	1172
Score:	715.00	Matches:	141
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0



QY 41 AsnGluLeuLeuGlnIlyLeuAspProLeuGlnGlnAlaIyValaIaPheValaSerAla 60  
Db 238 AATGAGTTGTCGAAAGTTGGATCCATCGAACAGCAAAAGGAGATTGGTTTCGCA 297  
QY 61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValaIaenProLyGlu 80  
Db 298 TACACATTAATTCAGTTTGGTTTATTGGCAACCCAGAGATTAACTCTAAGGAA 357  
QY 81 HisProValIyGlnGlnIlyLeuGlnIyArgIleArgValTyrMetAsnArgValIyGlnIle 100  
Db 358 CATCCAGTAAACAGCAATTGAAAGATCAGAGTATATATGACAGATCAAGCAAGAAATA 417  
QY 101 ThrAspIyValIyValaIyValaIyLeuAspArgGlyValaIaIaSerArgPheValaIyAsn 120  
Db 418 ACAGACAGAAAGAAAGCTGGCAGCTGACAGAGCTCAGCTTCAAGATTGTAAGAAAT 477  
QY 121 AlaLeuTrpGluProLySerLyAsnAlaSerLyValaIaIaenLyGlyLySerLy 140  
Db 478 GCCTCTGGGACCAAAATCGAAATGCAATCAAAAGTTGCCATTAAGGAAAGTAA 537  
QY 141 Ser 141  
Db 538 AGT 540  
RESULT 7  
LOCUS HSCIDPROT 1172 bp mRNA linear PRI 20-FEB-1998  
DEFINITION H.sapiens mRNA for CID protein.  
ACCESSION X95592  
VERSION X95592.1 GI:1185118  
KEYWORDS CID gene; CID protein; DNA-binding protein.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Nehle, P., Keck, T., Greferath, R., Spiess, E., Glaeser, T.,  
Rothbarth, K., Stammer, H. and Werner, D.  
TITLE cDNA cloning, recombinant expression and characterization of  
JOURNAL polypeptides with exceptional DNA affinity  
MEDLINE Nucleic Acids Res. 26 (5), 1160-1166 (1998)  
PUBMED 98136194  
REFERENCE  
MEDLINE 9469821  
PUBMED 2 (bases 1 to 1172)  
AUTHORS Werner, D.  
TITLE Direct Submission  
JOURNAL Submitted (08-FEB-1996) D. Werner, Dc. Krebsforschungszentrum,  
Biochemistry of the Cell, Im Neuenheimer Feld 280-0225, D-69120  
Heidelberg, FRG  
FEATURES  
source  
1. 1172  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CID-4"  
/tissue\_type="term placenta"  
/clone\_lib="lambda gtl0 (OG1)"  
1. 1172  
/gene="CID"  
118. 543  
/gene="CID"  
/note="DNA-binding protein"  
/codon\_start=1  
/product="CID protein"  
/protein\_id="CA64845.1"  
/db\_xref="GI:1185119"  
/db\_xref="GCA:Q13901"  
/db\_xref="EMBL:Q13901"  
/translation="MAGEEINDEYVPEIHETLSAFENSIGAVDEMLKTMASVSRNELL  
OKLDPLEQAKVDLVSAYTLNMFVYATGCVAPKPKHPVQQLERIRVYNNRVKEITLD  
KKVAGLDRGASRPFVKNMLWPKSKNASKVANKGSKS"  
polya\_signal  
1050. 1055

ORIGIN /gene="CID"  
Alignment Scores:  
Pred. No.: 6.88e-69 Length: 1172  
Score: 715.00 Matches: 141  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-701-618a-2 (1-141) x HSCIDPROT (1-1172)  
QY 1 MetAlaGlyGlnGlnIleAsnGlyAspTyrProValGlnIleHisGlyTyrLeuSerAla 20  
Db 118 ATGGCAGTAAAGAAATTAATGAAAGCTATCCAGTAAATTCACGAGTATTGTTCAGCG 177  
QY 21 PheGluAsnSerIleGlyValaIaValaIaPheGluMetLeuGlnThrMetSerValSerArg 40  
Db 178 TTTCAGAAATTCATGTTGCTGCTGAGATGAGATGCTGAAGACCATGATGCTGTTCTAGA 237  
QY 41 AsnGluLeuLeuGlnIlyLeuAspProLeuGlnGlnAlaIyValaIaPheValaSerAla 60  
Db 238 AATGAGTTGTCGAAAGTTGGATCCATCGAACAGCAAAAGGAGATTGGTTTCGCA 297  
QY 61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValaIaenProLyGlu 80  
Db 298 TACACATTAATTCAGTTTGGTTTATTGGCAACCCAGAGATTAACTCTAAGGAA 357  
QY 81 HisProValIyGlnGlnIlyLeuGlnIyArgIleArgValTyrMetAsnArgValIyGlnIle 100  
Db 358 CATCCAGTAAACAGCAATTGAAAGATCAGAGTATATATGACAGATCAAGCAAGAAATA 417  
QY 101 ThrAspIyValIyValaIyValaIyLeuAspArgGlyValaIaIaSerArgPheValaIyAsn 120  
Db 418 ACAGACAGAAAGAAAGCTGGCAGCTGACAGAGCTCAGCTTCAAGATTGTAAGAAAT 477  
QY 121 AlaLeuTrpGluProLySerLyAsnAlaSerLyValaIaIaenLyGlyLySerLy 140  
Db 478 GCCTCTGGGACCAAAATCGAAATGCAATCAAAAGTTGCCATTAAGGAAAGTAA 537  
QY 141 Ser 141  
Db 538 AGT 540  
RESULT 8  
LOCUS BC009589 1186 bp mRNA linear PRI 29-JUN-2004  
DEFINITION Homo sapiens nuclear DNA-binding protein, transcript variant 2,  
MNC (cDNA clone MGC:14659 IMAGE:4102498), complete cds.  
ACCESSION BC009589  
VERSION BC009589.1 GI:16307025  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, C.M., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.P., Zeeberg, B., Burow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Wax, S.I., Wang, J., Heien, F.,  
Diatchenko, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Uedln, T.B., Toshlyuk, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullehy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiyk, S.W.,  
Vallalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Ketteman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,









misc\_feature /note="match: GSS: Em:AQ436650"  
6588. .7012  
/note="match: GSS: Em:AQ324035"  
10015. .10842  
repeat\_region /note="LIPAI4 repeat: matches 5314. .6149 of consensus"  
11148. .12383  
/note="LIPAI4 repeat: matches 4066. .5314 of consensus"  
12901. .13700  
/note="LIPAI4 repeat: matches 8. .763 of consensus"  
17708. .18165  
/note="match: GSS: Em:AQ435928"  
18643. .19068  
/note="match: STS: Em:G52882  
match: GSS: Em:AQ202290"  
25943. .26064  
repeat\_region /note="LIPB3 repeat: matches 6026. .6147 of consensus"  
32371. .32545  
/note="LIM3C repeat: matches 1140. .1315 of consensus"  
36682. .38127  
/note="LIMB8 repeat: matches 4783. .6291 of consensus"  
38145. .42839  
/note="LIPB7 repeat: matches 1433. .6141 of consensus"  
43906. .44653  
/note="Cpg island"  
/evidence=not\_experimental  
44803. .46255  
/note="LIPB7 repeat: matches 9. .1460 of consensus"  
49657. .50116  
/note="Cpg island"  
/evidence=not\_experimental  
complement(54351. .55469)  
/gene="bAI9501.1"  
complement(54351. .55469)  
mRNA /product="bAI9501.1 (novel protein similar to CID  
DNA-binding protein)"  
/note="match: CDNAS: Em:X95592 Em:X95591  
match: ESTs: Em:AA831747 Em:T83118 Em:AA352554 Em:A1004660  
Em:AA442782 Em:AA305785 Em:A1804147 Em:A1809071  
Em:A1800846 Em:A151827 Em:AA805413 Em:AA835937 Em:NA0848  
Em:A1553826 Em:AA458588 Em:A1831757 Em:NA2131 Em:A1935349  
Em:N71224 Em:AM118590 Em:AM169841 Em:AA31129 Em:W87910  
Em:A1243111 Em:AA424253 Em:AM024957 Em:A1699498  
Em:A1898469 Em:A1310127 Em:AA418433 Em:A1123353  
Em:A1800532 Em:A1350721 Em:AA460755 Em:R68158 Em:AA861428  
Em:A1274315 Em:AA424184 Em:A1375251 Em:A1027628  
Em:A1301876 Em:AA476674 Em:A1816911 Em:A1476631  
Em:A1476155 Em:A1435863 Em:A1458192 Em:AM104208  
Em:AM340622 Em:A1379715 Em:A1671948 Em:W01341 Em:AM592440"  
/evidence=not\_experimental  
complement(54351)  
/gene="bAI9501.1"  
complement(54383. .54388)  
/gene="bAI9501.1"  
54386. .54689  
/note="match: STS: Em:G23952"  
complement(54399. .55406)  
/gene="bAI9501.1"  
/note="match: GSS: Em:AQ461100"  
complement(55051. .55386)  
/gene="bAI9501.1"  
/note="possible pseudogene  
match: proteins: Tr:Q13901"  
/codon\_start=1  
/evidence=not\_experimental  
/product="bAI9501.1 (novel protein similar to CID  
DNA-binding protein)"  
/protein\_id="CAP13204.1"  
/db\_xref="GI:17384063"  
/translation="MLKTVMSVRNELLQKLDLEQAKVDLSAYTLNSFWVYLATQ  
GVNPKHPIVKOELEIRIRVYMNVRVKEITDKKAKKLDKGAASRFKNALWEPKKNASK  
VANKKSKS"  
55412. .55487  
misc\_feature

repeat\_region /note="match: GSS: Em:AQ463761 Em:AQ505503"  
61512. .61611  
/note="50 copies 2 mer aa 61% conserved"  
68837. .68982  
repeat\_region /note="LIPAS repeat: matches 5997. .6143 of consensus"  
71617. .71998  
/note="LIPB3 repeat: matches 5769. .6150 of consensus"  
72363. .72506  
/note="72 copies 2 mer ta 77% conserved"  
78102. .78408  
/note="match: GSS: Em:AQ628574"  
complement(78524. .78950)  
/note="match: GSS: Em:AQ401519"  
78659. .78758  
repeat\_region /note="LIM3D repeat: matches 102. .199 of consensus"  
79868. .80242  
/note="LIPB8 repeat: matches 129. .503 of consensus"  
80250. .84310  
/note="LIPAI5 repeat: matches 2098. .6155 of consensus"  
84306. .84837  
/note="LIPAI5 repeat: matches 1407. .1939 of consensus"  
join(87084. .87121,87983. .88037,109294. .109356,  
111406. .111627,115595. .115629,118165. .118302,  
126770. .127043,127968. .128040,177990. .178100)  
/gene="bA479G22.1"  
join(87084. .87121,87983. .88037,109294. .109356,  
111406. .111627,115595. .115629,118165. .118302,  
126770. .127043,127968. .128040,177990. .178100)  
/gene="bA479G22.1"  
/product="bAI9501.3 (novel protein)"  
/note="isoform 2  
continues as bA453G10.1 in Em:AL356053  
match: CDNAS: Em:AK023093  
match: ESTs: Em:AA421130 Em:AA609088"  
/evidence=not\_experimental  
join(87084. .87121,87983. .88037,109294. .109356,  
111406. .111627,126770. .127043,127968. .128040,177990. .178100)  
/gene="bA479G22.1"  
/product="bAI9501.3 (novel protein)"  
/note="isoform 1  
continues as bA453G10.1 in Em:AL356053  
match: CDNAS: Em:AK023093  
match: ESTs: Em:AA421130"  
/evidence=not\_experimental  
complement(87095. .87660)  
/note="match: GSS: Em:AQ474468"  
88133. .88174  
/note="21 copies 2 mer tg 83% conserved"  
complement(88502. .88909)  
/note="match: GSS: Em:AQ129993"  
misc\_feature

Alignment Scores:  
Pred. No.: 2,33e-64 Length: 186856  
Score: 699.00 Matches: 138  
Percent Similarity: 98.57% Conservative: 0  
Best Local Similarity: 98.57% Mismatches: 2  
Query Match: 97.76% Indels: 0  
Gaps: 0  
DB: 9  
US-09-701-618a-2 (1-141) x AL354750 (1-186856)  
QY 2 AAGlGluGluIleAngluAePryrProValGluIleHisGluTyrlEuSerrAlaPhe 21  
DB 55473 GCAGGTGAAGAAATTAATGAAGACTATCCAGTGAATTCACAGATATTGTGCAGCCTT 55414  
QY 22 GUAASeTlGclGlyAlaValaIaepGluMeTleuTyThMeTetSeTValSerrArgAsn 41  
DB 55413 GAGAAATTCATTCGTCCTGCGATGAGATCGTGAACACACGATGCTGTTTCTAGAAAT 55354  
QY 42 GluLeuLeuGlnIuysEuAePProLeuGluGlnAlaIyevaIaIaPleuValSerrAlaTy 61  
DB 55353 GAATTGTCAGAAAGTTCGATTCACCTTGAACAAGCAAAAGTGGATTGTTCTGCATAC 55294



Db 98346 GCAGGTGAGAAATTAATGAAGCATCATGAGTAATTCAGATATTGTGACGGTTT 98287  
Qy 22 GUAAnSer1Ieg1yAlaValAspGluMetLeuLysThrMetCysValSerArgAsn 41  
Db 98286 GAGAAATTCATTTGCTGTGTGATGAGATGCTGAAGACCAAGATGCTGTTCTAGAAAT 98227  
Qy 42 GlnLeuLeuGlnLysLeuAspProLeuGlnGlnAlaLysValAspLeuValSerAlaTyr 61  
Db 98226 GAGTTGTCGACAAAGTTGATCCATTGACACAAAGATGAGTTGGTTTCTGCAATAC 98167  
Qy 62 ThrLeuAnSerMetPheTyrValTyrLeuAlaThrGlnGlyValAsnProLysGluHis 81  
Db 98166 ACATTAATTCATGTTTGGGTTTATTGGCAACCCAGAGATTATCTTAAGAAACAT 98107  
Qy 82 ProValLysGlnGlnLysLeuGlnLysGlnLeuValTyrMetAsnArgValLysGluIleThr 101  
Db 98106 CCAGTAAACAGCAAGATGGAAGATCATATATATGAAACAGATCAAGAAATACAA 98047  
Qy 102 AspLysLysLysValAspGluLysLeuAspArgGlyValAlaLysArgPheValLysAsnAla 121  
Db 98046 GACAAAGAAAGAGCTGGCAAGCTGACAGAGTGCAGCTTCAGATTGTAAAGAAATGCC 97987  
Qy 122 LeuTyrPgluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLysSer 141  
Db 97986 CTCTGGGACCAAAACCGAAATGATCATCAAAAGTTGCCATATAAGGAAAAAGTAAAGT 97927  
RESULT 13  
AX452900 426 bp DNA linear PAT 06-JUL-2002  
LOCUS AX452900  
DEFINITION Sequence 1 from Patent WO0242322.  
ACCESSION AX452900  
VERSION AX452900.1 GI:21712537  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 Jackson, D., Casari, G. and Suckow, J.  
AUTHORS Mammalian nuclear receptor cofactors c7 and c8 and methods of use  
TITLE Patent: WO 0242322-A 1 30-MAY-2002;  
JOURNAL LION Bioscience AG (DE)  
FEATURES  
source location/Qualifiers  
1..426  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.3e-64 Length: 426  
Score: 671.00 Matches: 132  
Percent Similarity: 95.74% Conservative: 3  
Best Local Similarity: 93.62% Mismatches: 6  
Query Match: 93.85% Indels: 0  
Gaps: 0  
DB: 6  
US-09-701-618a-2 (1-141) x AX452900 (1-426)  
Qy 1 MetAlaGlyGluGlnLysLeuAspTyrProValGlnIleHisGlyTyrLeuSerAla 20  
Db 1 ATGGGAGCTGAAGAAATTAATGAAGCATATCCAGTAATAATTCACGATTTATTTGTGCA 60  
Qy 21 PheGluAnSer1Ieg1yAlaValAspGluMetLeuLysThrMetCysValSerArg 40  
Db 61 TTGGCAATTCATGATGCTGTGATGAGATGCTGAAGAACATGATGCTGTTCTAGA 120  
Qy 41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnGlnAlaLysValAspLeuValSerAla 60  
Db 121 AATGAGTTGTGCAAGATGGAACCCATTGAACAAAGATGAGTTGGTTTCTGCA 180  
Qy 61 TyrThrLeuAnSerMetPheTyrValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80

Db 181 TACACATTAATTCATGTTTGGGTTTATTGGCAACTCAAGAGTGAATCTTAAGAA 240  
Qy 81 HisProValLysGlnGlnLysLeuGlnLysGlnLeuValTyrMetAsnArgValLysGluIle 100  
Db 241 CATCACTAAAGCAGAAATGGAAGATGAGATATATATGAAACAGATCAAGAAATA 300  
Qy 101 ThrAspLysLysLysValAspGluLysLeuAspArgGlyValAlaLysArgPheValLysAsn 120  
Db 301 ACAGACAAAGAAAGAGCTGGCAAGCTGACAGAGTGCAGCTTCAGATTGTGAAGAAAT 360  
Qy 121 AlaLeuTyrPgluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140  
Db 361 GCCCTTGGGACCAAAACCGAAATATCATCAAAAGTTGCCATTAAGGAAAAAGTAAA 420  
Qy 141 Ser 141  
Db 421 AGT 423  
RESULT 14  
AX452901/c 426 bp DNA linear PAT 06-JUL-2002  
LOCUS AX452901  
DEFINITION Sequence 2 from Patent WO0242322.  
ACCESSION AX452901  
VERSION AX452901.1 GI:21712538  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 Jackson, D., Casari, G. and Suckow, J.  
AUTHORS Mammalian nuclear receptor cofactors c7 and c8 and methods of use  
TITLE Patent: WO 0242322-A 2 30-MAY-2002;  
JOURNAL LION Bioscience AG (DE)  
FEATURES  
source location/Qualifiers  
1..426  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.3e-64 Length: 426  
Score: 671.00 Matches: 132  
Percent Similarity: 95.74% Conservative: 3  
Best Local Similarity: 93.62% Mismatches: 6  
Query Match: 93.85% Indels: 0  
Gaps: 0  
DB: 6  
US-09-701-618a-2 (1-141) x AX452901 (1-426)  
Qy 1 MetAlaGlyGluGlnLysLeuAspTyrProValGlnIleHisGlyTyrLeuSerAla 20  
Db 426 ATGGGAGCTGAAGAAATTAATGAAGCATATCCAGTAATAATTCACGATTTATTTGTGCA 367  
Qy 21 PheGluAnSer1Ieg1yAlaValAspGluMetLeuLysThrMetCysValSerArg 40  
Db 366 TTGGCAATTCATGATGCTGTGATGAGATGCTGAAGAACATGATGCTGTTCTAGA 307  
Qy 41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnGlnAlaLysValAspLeuValSerAla 60  
Db 306 AATGAGTTGTGCAAGATGGAACCCATTGAACAAAGATGAGTTGGTTTCTGCA 247  
Qy 61 TyrThrLeuAnSerMetPheTyrValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80  
Db 246 TACACATTAATTCATGTTTGGGTTTATTGGCAACTCAAGAGTGAATCTTAAGAA 187  
Qy 81 HisProValLysGlnGlnLysLeuGlnLysGlnLeuValTyrMetAsnArgValLysGluIle 100  
Db 186 CATCACTAAAGCAGAAATGGAAGATGAGATATATATGAAACAGATCAAGAAATA 127  
Qy 101 ThrAspLysLysLysValAspGluLysLeuAspArgGlyValAlaLysArgPheValLysAsn 120

Db	126	ACMGACAGAAAAAGGCTGCGCAAGCTGCGACAGAGGTCGACGCTTCAAGATTGTGAAGAAAT	67
Qy	121	AlaLeuTgPjUProLySeSerLySaAnaLaSeRtLySvAlaLaAsnLySgLySeSerLyS	140
Db	66	GCCCTCTGGAGACCAAAACCGAAAAATATACATCCAAAGTTGCCCATTAAGAAAAAGTAA	7
Qy	141	Set	141
Db	6	ACT	4
RESULT 15			
AL356095/c			
LOCUS			
DEFINITION	Human DNA sequence from clone RP11-369J21 on chromosome 10,		
ACCESSION	AL356095		
VERSION	AL356095.11	GI:12331029	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 172079)		
JOURNAL	Brown, A.		
COMMENT	Direct Submission Submitted (24-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jan 22, 2001 this sequence version replaced gi:11863400. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known ambiguated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone confis of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr10 RP11-369J21 is from the library RPCT-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VICTOR: PBAC3.6 This sequence is the entire insert of clone RP11-369J21 The true right end of clone RP11-479017 is at 65134 in this sequence. Location/Qualifiers 1. 172079 /organism="Homo sapiens" /mol_type="genomic DNA" /dd_xref="taxon:9606" /chromosome="10" /clone="RP11-369J21" /clone_1lb="RPCT-11.2" 7. 431 /note="match: GSS: Em:AQ191918" 25. 413 /note="match: GSS: Em:AQ815029" 61. 595 /note="match: GSS: Em:AQ531010" 647. 944 /note="11P1A12 repeat: matches -1422. -1118 of consensus" 1339. 11734 /note="L2 repeat: matches 1652. 2042 of consensus"		

repeat_region	2037. 2348	/note="AluXx repeat: matches 1. .312 of consensus"
repeat_region	2389. .2697	/note="AluXx repeat: matches 2. .310 of consensus"
repeat_region	2939. .3074	/note="FLAM_C repeat: matches 12. .141 of consensus"
repeat_region	3197. .3411	/note="MER58A repeat: matches 1. .224 of consensus"
repeat_region	3518. .3821	/note="AluXx repeat: matches 1. .300 of consensus"
repeat_region	4024. .4580	/note="L2 repeat: matches 2210. .2750 of consensus"
repeat_region	4626. .4766	/note="L2 repeat: matches 2240. .2393 of consensus"
repeat_region	4768. .5076	/note="AluXx repeat: matches 1. .312 of consensus"
repeat_region	5475. .5997	/note="L1MBD repeat: matches 102. .831 of consensus"
repeat_region	5994. .6157	/note="L1MB repeat: matches 4140. .4298 of consensus"
repeat_region	6158. .6294	/note="FLAM_C repeat: matches 1. .133 of consensus"
repeat_region	6295. .6750	/note="L1MB repeat: matches 4298. .4741 of consensus"
repeat_region	6751. .7063	/note="AluY repeat: matches 1. .311 of consensus"
repeat_region	7064. .7628	/note="L1MB repeat: matches 4741. .5310 of consensus"
repeat_region	7629. .7915	/note="AluXx repeat: matches 1. .284 of consensus"
repeat_region	7916. .8047	/note="L1MB repeat: matches 5310. .5437 of consensus"
repeat_region	8054. .8180	/note="FLAM_C repeat: matches 2. .140 of consensus"
repeat_region	8182. .8411	/note="L1PA7 repeat: matches 5432. .5661 of consensus"
repeat_region	8412. .8717	/note="AluXx repeat: matches 1. .306 of consensus"
repeat_region	8718. .8862	/note="L1PA7 repeat: matches 5661. .5805 of consensus"
repeat_region	8863. .9012	/note="L1PA7 repeat: matches 5992. .6141 of consensus"
repeat_region	9140. .9288	/note="AluDo repeat: matches 1. .151 of consensus"
repeat_region	9292. .9516	/note="L1MB6 repeat: matches 5971. .6182 of consensus"
repeat_region	10088. .10389	/note="AluXx repeat: matches 1. .302 of consensus"
repeat_region	10390. .10697	/note="AluY repeat: matches 1. .308 of consensus"
repeat_region	11104. .11317	/note="MER46A repeat: matches 22. .235 of consensus"
repeat_region	11749. .12047	/note="AluXx repeat: matches 1. .299 of consensus"
repeat_region	12418. .12472	/note="L1MB2 repeat: matches 3968. .4023 of consensus"
repeat_region	12473. .12772	/note="AluXx repeat: matches 1. .303 of consensus"
repeat_region	12773. .13104	/note="L1MB2 repeat: matches 4023. .4362 of consensus"
repeat_region	13105. .13408	/note="AluXx repeat: matches 1. .300 of consensus"
repeat_region	13409. .13722	/note="AluY repeat: matches 1. .310 of consensus"
repeat_region	13410. .14039	/note="L1MB2 repeat: matches 4362. .4638 of consensus"
repeat_region	14040. .14344	/note="AluXx repeat: matches 1. .300 of consensus"
repeat_region	14345. .14683	/note="L1MB2 repeat: matches 4638. .4985 of consensus"
repeat_region	14745. .15005	/note="AluXx repeat: matches 1. .291 of consensus"
misc feature	complement(14996. .15536)	



**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 4, 2005, 03:16:16 : Search time 5271.49 Seconds  
(without alignments)  
10370.300 Million cell updates/sec

Title: US-09-701-618A-1

Perfect score: 1156  
Sequence: 1 cttccggagagactggagtc.....aggctgaataatgttgaataaa 1156

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenBdb1:\*  
1: gb\_ba:\*  
2: gb\_ptg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1156	100.0	1156	6	BD221634
2	1156	100.0	1156	6	AX009361
3	1145	99.0	1172	6	AX328809
4	1145	99.0	1172	6	AX779798
5	1145	99.0	1172	6	AX779799
6	1145	99.0	1172	6	HSCIDPRROT
7	927.2	80.2	1168	9	BC009589
8	927.2	80.1	1168	9	BC016284
9	922.4	79.8	1195	9	BC009584
10	919.8	77.9	1199	9	BC005235
11	900.4	77.9	186856	2	AC068741
12	900.4	77.9	195426	2	AC068741
13	655.8	56.7	155888	9	AL356095
14	519.2	44.9	172079	9	AL356095
15	464	40.1	2000	10	BC005436
16	458	39.6	1040	10	AF031426
17	433.8	37.5	1040	6	BD221635
18	433.8	37.5	1040	6	AX009363
19	433.8	37.5	1050	10	MWCIDPRROT

20	431	37.3	443	6	AX777852
21	414.2	35.8	215616	2	AC109025
22	414.2	35.8	224919	2	AC110304
23	400.4	34.6	426	6	AX452900
24	400.4	34.6	426	6	AX452901
25	394.4	34.1	420	6	AX467454
26	394.4	34.1	420	6	AX467455
27	389.6	33.7	420	6	AX452903
28	389.6	33.7	420	6	AX452904
29	389.2	33.7	789	10	AY302220
30	376.8	32.6	457	6	CQ704080
31	312	27.0	380	6	AX885372
32	312	27.0	380	6	BD024982
33	296.2	25.6	154312	9	AP001767
34	286.2	25.6	178728	2	AC087613
35	285.6	23.0	212115	2	AC116701
36	285.6	23.0	238058	2	AC114654
37	265.2	22.9	395	11	G23952
38	259.4	22.4	342	11	BV179135
39	239.4	20.7	232869	10	AL603925
40	239.4	20.7	235365	2	AC044845
41	239	20.7	4018	9	HSW807917
42	200.4	17.3	728	5	CR353144
43	194.8	16.9	209355	2	AC123163
44	194.8	16.9	221055	2	AC097131
45	154.4	13.4	700	11	BV072361

## ALIGNMENTS

RESULT 1	BD221634	1156 bp	DNA	linear	PAT 17-JUN-2003
LOCUS	BD221634	Method of inducing apoptosis in cell.			
DEFINITION	BD221634.1	GI:33031404			
ACCESSION	BD221634.1	GI:33031404			
VERSION	JP 2002517192-A/1				
KEYWORDS	JP 2002517192-A/1				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 1156)				
AUTHORS	Rothbarth, K., Stammer, H., Werner, D. and Nails, P.				
TITLE	Method of inducing apoptosis in cell				
JOURNAL	Patent: JP 2002517192-A 1 18-JUN-2002;				
DEUTSCHES KREBSFORSCHUNGSZENTRUM STIFTUNG DES OFFENTLICHEN RECHTS,					
PETER NAILS					
COMMENT	OS Homo sapiens (human)				
PN	JP 2002517192-A/1				
PD	18-JUN-2002				
PF	03-JUN-1999 JP 2000552267				
PR	03-JUN-1998 DE 198 24 811.3				
PI	KARSTEN ROTHBARTH, HERMANN STAMMER, DIETER WERNER, PETER NAILS PC				
C12N15/09	A61K38/00, A61K48/00, A61P35/00, C12N15/00, A61K37/02 CC				
FEATURES	Method of inducing apoptosis in cell				
source	FT CDS (118) (540)				
FT	mat_peptide (118) (540)				
location/Qualifiers	1. 1156				
location/Qualifiers	/organism="Homo sapiens"				
location/Qualifiers	/mol_type="genomic DNA"				
location/Qualifiers	/db_xref="taxon:9606"				
ORIGIN					
Query Match	100.0%	Score 1156	DB 6	Length 1156	
Best Local Similarity	100.0%	Pred. No. 6.3e-197	Indels 0	Gaps 0	
Matches 1156	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
1	CTTCCGGAGAGCTGAGTGAAGCCGCTGAGTATTTCTAAGCCAGTGTGAGAGTA	60			
1	CTTCCGGAGAGCTGAGTGAAGCCGCTGAGTATTTCTAAGCCAGTGTGAGAGTA	60			
Db	1	CTTCCGGAGAGCTGAGTGAAGCCGCTGAGTATTTCTAAGCCAGTGTGAGAGTA	60		

QY 61 TGTGAGGCAAGTACCTATAGAAACCGGAGAGGGTGAGAGACAGCTGGCCATATG 120  
 DB 61 TGTGAGGCAAGTACCTATAGAAACCGGAGAGGGTGAGAGACAGCTGGCCATATG 120  
 QY 121 GCAGGTGAGAAATTAATGAAAGACTATCCAGTAGAATTCACGATATTGTGACGCTTT 180  
 DB 121 GCAGGTGAGAAATTAATGAAAGACTATCCAGTAGAATTCACGATATTGTGACGCTTT 180  
 QY 181 GAGAAATTCATTTGGTGTGTGAGATGAGATGCTGAAAGACATGATGTCTGTTCTTGAAAT 240  
 DB 181 GAGAAATTCATTTGGTGTGTGAGATGAGATGCTGAAAGACATGATGTCTGTTCTTGAAAT 240  
 QY 241 GAGTTGTGACAGAGTTGATCCACTTGAACAGCAAAAGTGATTTGGTTCTGCAATAC 300  
 DB 241 GAGTTGTGACAGAGTTGATCCACTTGAACAGCAAAAGTGATTTGGTTCTGCAATAC 300  
 QY 301 ACATTAATTCATTTGGGTTTATTTGGCAACCCAGAGATTAACTCTTAGGAACAT 360  
 DB 301 ACATTAATTCATTTGGGTTTATTTGGCAACCCAGAGATTAACTCTTAGGAACAT 360  
 QY 361 CCAGTAAACAGGAATTGGAAGAATCAGAGTATATGAAACAGATCAAGGAATTAACA 420  
 DB 361 CCAGTAAACAGGAATTGGAAGAATCAGAGTATATGAAACAGATCAAGGAATTAACA 420  
 QY 421 GACAAGAAAAAGCTGGCAAGCTGGAACAGAGGTGAGCTTCAAGATTGTAAAAAATGCC 480  
 DB 421 GACAAGAAAAAGCTGGCAAGCTGGAACAGAGGTGAGCTTCAAGATTGTAAAAAATGCC 480  
 QY 481 CTCTGGGAACCAAAATGCAAAATGCAAAAGTTGCCATTAAGAAAAAGTAAAAAGT 540  
 DB 481 CTCTGGGAACCAAAATGCAAAATGCAAAAGTTGCCATTAAGAAAAAGTAAAAAGT 540  
 QY 541 TAACTTTTGGTTTGTGATGACATATTCAAAAGTCAATTAATGTATATCAAGTAA 600  
 DB 541 TAACTTTTGGTTTGTGATGACATATTCAAAAGTCAATTAATGTATATCAAGTAA 600  
 QY 601 TATGTAAGCTAAATACCTCTCCAAAGATCAATTAATCTTATGATTAGACCTGAGGA 660  
 DB 601 TATGTAAGCTAAATACCTCTCCAAAGATCAATTAATCTTATGATTAGACCTGAGGA 660  
 QY 661 TTTTAAATGTGATATATATATATATATATTAATTAACATCTCTGATGAGACCTTAT 720  
 DB 661 TTTTAAATGTGATATATATATATATATATTAATTAACATCTCTGATGAGACCTTAT 720  
 QY 721 TCTTAAATAGTGAAGCTTGCAGAGTACCTTTTAAAGCAGCTGGAATTTAAGTAA 780  
 DB 721 TCTTAAATAGTGAAGCTTGCAGAGTACCTTTTAAAGCAGCTGGAATTTAAGTAA 780  
 QY 781 ATGTTCTTTGTAACATTTGTAATTTTAAATGAATGAACCTTATGAAATGCTAT 840  
 DB 781 ATGTTCTTTGTAACATTTGTAATTTTAAATGAATGAACCTTATGAAATGCTAT 840  
 QY 841 CTGAGGCTGAATTAATAGTACATCTGTTTCACTATATGATATTAAGAAAGCGTAA 900  
 DB 841 CTGAGGCTGAATTAATAGTACATCTGTTTCACTATATGATATTAAGAAAGCGTAA 900  
 QY 901 GACTTAATGTCAATTTTCTGATAGATCTTAAATCAATGTTTCAATGATTTTAAAGAA 960  
 DB 901 GACTTAATGTCAATTTTCTGATAGATCTTAAATCAATGTTTCAATGATTTTAAAGAA 960  
 QY 961 TTAAGCTTTGTTGATATTCAGAGTGAACCTAAAGTTTAAAGTTTAAATTTCT 1020  
 DB 961 TTAAGCTTTGTTGATATTCAGAGTGAACCTAAAGTTTAAAGTTTAAATTTCT 1020  
 QY 1021 TGGCATTTGCTCTATGCTCCATTTAAATAAATCAATTTCTATTAACCTTAGATGG 1080  
 DB 1021 TGGCATTTGCTCTATGCTCCATTTAAATAAATCAATTTCTATTAACCTTAGATGG 1080  
 QY 1081 AAATAAGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
 DB 1081 AAATAAGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140

QY 1141 TGAATATGTTGTAAAA 1156  
 DB 1141 TGAATATGTTGTAAAA 1156  
 RESULT 2  
 LOCUS AX009361  
 DEFINITION Sequence 1 from Patent WO9963071.  
 ACCESSION AX009361  
 VERSION AX009361.1 GI:9996662  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1  
 AUTHORS Rothbarth,K., Werner,D. and Stammer,H.  
 TITLE Method for triggering apoptosis in cells  
 JOURNAL Patent: WO 9963071-A1 09-DEC-1999;  
 DEUTSCHES KREBSFORSCH (DE); ROTHBARTH KARSTEN (DE); WERNER DIETER (DE); STAMMER HERMANN (DE)  
 FEATURES  
 source  
 location/Qualifiers  
 1..1156  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 118..543  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC07450.1"  
 /db\_xref="GI:9996663"  
 /translation="MAGEKINEDYPIEIHLYLSAFENSIGAVDEMLKTMMSVSRNEL  
 QKIDPQBQKVDIVSAVTTLSMFWLATDGVAPKHPVQELERIRVIMRVKEITD  
 KKRAAGKLDRGAASRFVKNLWEPKSNASRVANKRSKS"  
 ORIGIN  
 Query Match 100.0%; Score 1156; DB 6; Length 1156;  
 Best Local Similarity 100.0%; Pred. No. 6, 3e-197; Indels 0; Gaps 0;  
 Matches 1156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTTTCCGGGAGACTGAGTGAAGGCGTGAGTATTTCTTAAGCCAGTGTGAGAGTA 60  
 DB 1 CTTTCCGGGAGACTGAGTGAAGGCGTGAGTATTTCTTAAGCCAGTGTGAGAGTA 60  
 QY 61 TGTGAGGCAAGATCCTATAGAAACCGGAGAGGGTGAGAGACAGAGCTGGCCATATG 120  
 DB 61 TGTGAGGCAAGATCCTATAGAAACCGGAGAGGGTGAGAGACAGAGCTGGCCATATG 120  
 QY 121 GCAGGTGAGAAATTAATGAAAGACTATCCAGTAGAATTCACGATATTGTGACGCTTT 180  
 DB 121 GCAGGTGAGAAATTAATGAAAGACTATCCAGTAGAATTCACGATATTGTGACGCTTT 180  
 QY 181 GAGAAATTCATTTGGTGTGTGAGATGAGATGCTGAAAGACATGATGTCTGTTCTTGAAAT 240  
 DB 181 GAGAAATTCATTTGGTGTGTGAGATGAGATGCTGAAAGACATGATGTCTGTTCTTGAAAT 240  
 QY 241 GAGTTGTGACAGAGTTGATCCACTTGAACAGCAAAAGTGATTTGGTTCTGCAATAC 300  
 DB 241 GAGTTGTGACAGAGTTGATCCACTTGAACAGCAAAAGTGATTTGGTTCTGCAATAC 300  
 QY 301 ACATTAATTCATTTGGGTTTATTTGGCAACCCAGAGATTAACTCTTAGGAACAT 360  
 DB 301 ACATTAATTCATTTGGGTTTATTTGGCAACCCAGAGATTAACTCTTAGGAACAT 360  
 QY 361 CCAGTAAACAGGAATTGGAAGAATCAGAGTATATGAAACAGATCAAGGAATTAACA 420  
 DB 361 CCAGTAAACAGGAATTGGAAGAATCAGAGTATATGAAACAGATCAAGGAATTAACA 420  
 QY 421 GACAAGAAAAAGCTGGCAAGCTGGAACAGAGGTGAGCTTCAAGATTGTAAAAAATGCC 480  
 DB 421 GACAAGAAAAAGCTGGCAAGCTGGAACAGAGGTGAGCTTCAAGATTGTAAAAAATGCC 480

QY 481 CTCGGGACCAAAATGCAAAATGATCAAAAGTTGCCAATTAAGGAAAAAGTAAAGT 540  
Db 481 CTCGGGACCAAAATGCAAAATGATCAAAAGTTGCCAATTAAGGAAAAAGTAAAGT 540  
QY 541 TAACTTTTGGTTTGTGATGACATATTCAAAAGTACATTAATATGTAATCAGACTAA 600  
Db 541 TAACTTTTGGTTTGTGATGACATATTCAAAAGTACATTAATATGTAATCAGACTAA 600  
QY 601 TATGTAAAGTAAATCTCTCTCCCAAAATCATTTATTTATGTAATGACACTGAGA 660  
Db 601 TATGTAAAGTAAATCTCTCTCCCAAAATCATTTATTTATGTAATGACACTGAGA 660  
QY 661 TTTTAACTTGTATATATATATATATTAATTAATTAATTAATTAATTAATTAATTA 720  
Db 661 TTTTAACTTGTATATATATATATATTAATTAATTAATTAATTAATTAATTAATTA 720  
QY 721 TCTTTATATAGTCACTTGTGCAAGTACATTTTATTAAGAGCTGTGAAAATTAAGTAA 780  
Db 721 TCTTTATATAGTCACTTGTGCAAGTACATTTTATTAAGAGCTGTGAAAATTAAGTAA 780  
QY 781 ATGTTCTTTGTAAACATTTGTACTATTTTAAATGAATGAATGAATGAATGAATGAAT 840  
Db 781 ATGTTCTTTGTAAACATTTGTACTATTTTAAATGAATGAATGAATGAATGAATGAAT 840  
QY 841 CTGTAGGCTGAAATTAATAGTACATCTGTTTCACTATATATTAATTAAGAGCTGAAAT 900  
Db 841 CTGTAGGCTGAAATTAATAGTACATCTGTTTCACTATATATTAATTAAGAGCTGAAAT 900  
QY 901 GACTTAAATGTTCAATTTTCTGTATATGATTAATTAATTAATTAATTAATTAATTA 960  
Db 901 GACTTAAATGTTCAATTTTCTGTATATGATTAATTAATTAATTAATTAATTAATTA 960  
QY 961 TTAATCTTTGTGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020  
Db 961 TTAATCTTTGTGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020  
QY 1021 TGGCATGTTGCTCTATATGCTCCATTTTAAATTAATTAATTAATTAATTAATTAAT 1080  
Db 1021 TGGCATGTTGCTCTATATGCTCCATTTTAAATTAATTAATTAATTAATTAATTAAT 1080  
QY 1081 AATTAAGTGTATATGTTGATGATGAATTTTGGCATGATGATGATGATGATGATGAT 1140  
Db 1081 AATTAAGTGTATATGTTGATGATGAATTTTGGCATGATGATGATGATGATGATGAT 1140  
QY 1141 TGAATAATGTTGTAATA 1156  
Db 1141 TGAATAATGTTGTAATA 1156

RESULT 3  
AX329809 1172 bp DNA linear PAT 09-JAN-2002  
LOCUS AX329809  
DEFINITION Sequence 318 from Patent WO0194629.  
ACCESSION AX329809  
VERSION AX329809.1 GI:18102787  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,  
Horizigan, S., Soppet, D.R. and Weaver, Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
gene sets  
JOURNAL Patent: WO 0194629-A 318 13-DEC-2001;  
Avalon Pharmaceuticals (US)  
FEATURES  
source location/Qualifiers  
1..1172  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 99.9%; Score 1145; DB 6; Length 1172;  
Best Local Similarity 99.9%; Pred. No. 5.8e-195;  
Matches 1156; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1 CTTTCCGGAGAGCTGAGAGTGAAGAGCCGTGATATTTCTTAAGCCAGTGTTAAGAGTA 60  
Db 1 CTTTCCGGAGAGCTGAGAGTGAAGAGCCGTGATATTTCTTAAGCCAGTGTTAAGAGTA 60  
QY 61 TGTGAGGCAAGAGTACCTATTAAGAACCCGAGAGAGGGTGAAGAGCAGAGCTGGCATATG 120  
Db 61 TGTGAGGCAAGAGTACCTATTAAGAACCCGAGAGAGGGTGAAGAGCAGAGCTGGCATATG 120  
QY 121 GCAAGTGAAGAAATTAATGAAGACTATCCAGTGAATTCACAGATTTTGTACAGCTTT 180  
Db 121 GCAAGTGAAGAAATTAATGAAGACTATCCAGTGAATTCACAGATTTTGTACAGCTTT 180  
QY 181 GAGAAATTCATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
Db 181 GAGAAATTCATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
QY 241 GAGTTGTGCAAGATTGATCACTTGAACAGCAAAAGTGAATTTGTTTGTGCAATAC 300  
Db 241 GAGTTGTGCAAGATTGATCACTTGAACAGCAAAAGTGAATTTGTTTGTGCAATAC 300  
QY 301 ACATTAATTCATATGTTTGGGTTTATTTGGCAACCCAGAGTTAATCTTAAGAAATCAT 360  
Db 301 ACATTAATTCATATGTTTGGGTTTATTTGGCAACCCAGAGTTAATCTTAAGAAATCAT 360  
QY 361 CCAGTAAACAGAGAAATTTGAAGAAATCAGAGTATATATGAACAGAGTCAAGAAATTAACA 420  
Db 361 CCAGTAAACAGAGAAATTTGAAGAAATCAGAGTATATATGAACAGAGTCAAGAAATTAACA 420  
QY 421 GACAGAAAAAGGCTGGCAAGCTGGAACAGAGTGCAGCTTCAAGATTTGTAAAAATGCC 480  
Db 421 GACAGAAAAAGGCTGGCAAGCTGGAACAGAGTGCAGCTTCAAGATTTGTAAAAATGCC 480  
QY 481 CTCGGGACCAAAATGCAAAATGATCAAAAGTTGCCAATTAAGGAAAAAGTAAAGT 540  
Db 481 CTCGGGACCAAAATGCAAAATGATCAAAAGTTGCCAATTAAGGAAAAAGTAAAGT 540  
QY 541 TAACTTTTGGTTTGTGATGACATATTCAAAAGTACATTAATATGTAATCAGACTAA 600  
Db 541 TAACTTTTGGTTTGTGATGACATATTCAAAAGTACATTAATATGTAATCAGACTAA 600  
QY 601 TATGTAAAGTAAATCTCTCTCCCAAAAGTACATTAATTAATTAATTAATTAATTAAT 660  
Db 601 TATGTAAAGTAAATCTCTCTCCCAAAAGTACATTAATTAATTAATTAATTAATTAAT 660  
QY 661 TTTTAACTTGTATATATATATATATTAATTAATTAATTAATTAATTAATTAATTAAT 720  
Db 661 TTTTAACTTGTATATATATATATATTAATTAATTAATTAATTAATTAATTAATTAAT 720  
QY 721 TCTTTATATAGTCACTTGTGCAAGTACATTTTATTAAGAGCTGTGAAAATTAAGTAA 780  
Db 721 TCTTTATATAGTCACTTGTGCAAGTACATTTTATTAAGAGCTGTGAAAATTAAGTAA 780  
QY 781 ATGTTCTTTGTAAACATTTGTACTATTTTAAATGAATTAATGAATTAATGAATTAAT 840  
Db 781 ATGTTCTTTGTAAACATTTGTACTATTTTAAATGAATTAATGAATTAATGAATTAAT 840  
QY 841 CTGTAGGCTGAAATTAATAGTACATCTGTTTCACTATATGATATTAAGAACGTCG-AA 899  
Db 841 CTGTAGGCTGAAATTAATAGTACATCTGTTTCACTATATGATATTAAGAACGTCGAAA 900  
QY 900 TGACTTAATATGTTTCTATATATATATATATATATATATATATATATATATATATAT 959  
Db 901 TGACTTAATATGTTTCTATATATATATATATATATATATATATATATATATATATAT 960  
QY 960 ATTACGCTTTGTGATATTAATCAAGTGTGAACCTAAAGTTAATGTTGATCTTAATTC 1019  
Db 961 ATTACGCTTTGTGATATTAATCAAGTGTGAACCTAAAGTTAATGTTGATCTTAATTC 1020

Qy	1020	TTGGCAGTGTGCTCTATGTCCATTAAATAAATACATTCCTCAATTAAGTTGATGG	1079
Db	1021	TTGGCAGTGTGCTCTATGTCCATTAAATAAATACATTCCTCAATTAAGTTGATGG	1080
Qy	1080	GAATTAAGGTTGATGTGATGATGAATTTTGGCATGATGACGTCACTCATTAAGG	1139
Db	1081	GAATTAAGGTTGATGTGATGATGAATTTTGGCATGATGACGTCACTCATTAAGG	1140
Qy	1140	CTGAATAATGTTGTAAAA	1156
Db	1141	CTGAATAATGTTGTAAAA	1157

RESULT 4		
AX779798		
LOCUS	AX779798	1172 bp DNA
DEFINITION	Sequence 1955 from Patent WO0303443.	linear
ACCESSION	AX779798	
VERSION	AX779798.1	GI:32696792
KEYWORDS	.	
SOURCE	Homo sapiens (human)	

ORGANISM  
Homo sapiens (Human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini, Homnidae; Homo.  
REFERENCE  
1

**TITLE** Novel genetic markers for leukemias  
**JOURNAL** Patent: WO 03039443-A 1955 15-MAY-2003.

**FEATURES**

Ludwig-Maximilians-Universität München (DE) ; Häferlach, Torsten  
PD Dr. (DE) ; Schoch, Claudia (DE) ; Kern, Wolfgang (DE)  
Location/Qualifiers

```

/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

```

Query Match	99.0%	Score 1145;	DB 6;	Length 1172;
Best Local Similarity	99.9%	Pred. No. 5.8e-195;		
Matches 1156; Conservative	0;	Mismatches	0;	Indels 1; Gaps 1

QY	1	CTTCCGGGAGACTGAGATCGAAGCCGCTGAGATATTTCTAAGCCAGTGTATTGAGAGTA	60
Db	1	CTTCCGGGAGACTGAGATCGAAGCCGCTGAGTAATTTCTAAGCCAGTGTATTGAGAGTA	60
QY	61	TGTGAGCGAAGAGTACCCTATAGAACCCGGAGAGAGGTGAGGACGAGAGCTGGCCATTAATG	120
Db	61	TGTGAGCGAAGAGTACCTATAGAACCCGGAGAGAGGTGAGGACGAGAGCTGGCCATTAATG	120
QY	121	GCAGGTGAAGAAATTAATGAAAGTACATATCCAGTAAATTCAGAGATATTTGACGCGTTT	180
Db	121	GCAGGTGAAGAAATTAATGAAAGTACATATCCAGTAAATTCAGAGATATTTGACGCGTTT	180
QY	181	GAGAAATTCATATGCTGCTGTGATGATGATGCTGAAGACCATGATGTCGTTTCTAGAAAT	240
Db	181	GAGAAATTCATATGCTGCTGTGATGATGATGCTGAAGACCATGATGTCGTTTCTAGAAAT	240
QY	241	GAGTGTGTGCAAGAGTTGATGCATCCCTTGAAACAACCAAAAGTGATTTGGTTTCTGCATAC	300
Db	241	GAGTGTGTGCAAGAGTTGATGCATCCCTTGAAACAACCAAAAGTGATTTGGTTTCTGCATAC	300
QY	301	ACATTAAATTCATGTATTTGGGTTTATTTTGCAACCCAAAGAGTTAATCCCTAAGGAACAT	360
Db	301	ACATTAAATTCATGTATTTGGGTTTATTTTGCAACCCAAAGAGTTAATCCCTAAGGAACAT	360
QY	361	CCAGTAAACACGGAATTCGAAAGATCAGAGTATATATGAACAGATCAAGCAAAATPACA	420
Db	361	CCAGTAAACACGGAATTCGAAAGATCAGAGTATATATGAACAGATCAAGCAAAATPACA	420
QY	421	GACAAAGAAAAGCTGGCAAGCTGGAAGAGGTGCAGCTTCAAGATTTGTAAAAATGCC	480

Db	421	GACAAAGAAAAGGCGCTGGCAAGCTGGACAGAGGTCAGCTTCAAGATTGTAAAAAATGCC	480
Qy	481	CTTGGGAGACCCAAATATCGAAAAATATGCATCAAAAGTTGCGCAATTAAGGAAAAAGTTAAAGT	540
Db	481	CTTGGGAGACCCAAATATCGAAAAATATGCATCAAAAGTTGCGCAATTAAGGAAAAAGTTAAAGT	540
Qy	541	TAACTTTTGGTTTTGATGTATACATATTCAAAAAGTACATTAATATGTAATCACAGTAA	600
Db	541	TAACTTTTGGTTTTGATGTATACATATTCAAAAAGTACATTAATATGTAATCACAGTAA	600
Qy	601	TATGTAAAGCTAAATATACCTTCCTCCCAAGATCATATCTTTATGTAATAGACCTGAGGA	660
Db	601	TATGTAAAGCTAAATATACCTTCCTCCCAAGATCATATCTTTATGTAATAGACCTGAGGA	660
Qy	661	TTTTAAACATGTGATATATATATATATTTATATTTTCCATCTCTTGATGAGACTCTTAT	720
Db	661	TTTTAAACATGTGATATATATATATATTTATATTTTCCATCTCTTGATGAGACTCTTAT	720
Qy	721	TCTTTATATAGTCAAGCTCTTGCAGAGTACCAATTTTAAAGCAGCTGGAATTAAGTAA	780
Db	721	TCTTTATATAGTCAAGCTCTTGCAGAGTACCAATTTTAAAGCAGCTGGAATTTAAGTAA	780
Qy	781	ATGTCCTTGGTAAACATTTGTACTATTTTAAATGAATAATGACCTTATGAAGTATGCTAT	840
Db	781	ATGTCCTTGGTAAACATTTGTACTATTTTAAATGAATAATGACCTTATGAAGTATGCTAT	840
Qy	841	CTGTAGAGCTAAATTAAGTACATCTGTGTTTCACTATATGATATTAAGAAAGGCTG-AA	899
Db	841	CTGTAGAGCTAAATTAAGTACATCTGTGTTTCACTATATGATATTAAGAAAGGCTG-AA	899
Qy	900	TGACTTAATATGTCATTTTTTCTGTATAGATCTTATATCATATGTTTCATGATTTTAAAG	959
Db	901	TGACTTAATATGTCATTTTTTCTGTATAGATCTTATATCATATGTTTCATGATTTTAAAG	960
Qy	960	ATTACTGCTTTGTATATATCAAGATGGAAGCACTAAAGCTTAATAGTGTACTTAATTC	1019
Db	961	ATTACTGCTTTGTATATATCAAGATGGAAGCACTAAAGCTTAATAGTGTACTTAATTC	1020
Qy	1020	TTCGCAATGTGCTTATATGTCCTCAATTTAAATTAATATACATTCCTCAATTAAGTATG	1079
Db	1021	TTCGCAATGTGCTTATATGTCCTCAATTTAAATTAATTAATATACATTCCTCAATTAAGT	1080
Qy	1080	GAAATTAAGTTGATATGATGATGAAATTTTGCATGATGAGACGTACTCTCAATAAGG	1139
Db	1081	GAAATTAAGTTGATATGATGATGAAATTTTGCATGATGAGACGTACTCTCAATAAGG	1140
Qy	1140	CTGAAAAATGTTGTAAAA	1156
Db	1141	CTGAAAAATGTTGTAAAA	1157

FEATURES	source	1172 bp	DNA	linear	PAT 14-JUL-2003
LOCUS	AX779799				
DEFINITION	Sequence 1956 from Patent WO0309443.				
ACCESSION	AX779799				
VERSION	AX779799.1	GI:32696793			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1				
JOURNAL	Hafeljach, T., Schoch, C., Kern, W., Kohlmann, A., Schittiger, S., Dugas, M., Ellis, R., Bross, B. and Mergenthaler, S. Novel genetic markers for leukemias Patent: WO 03039443-A 1996 15-MAY-2003; Deutsches Krebsforschungszentrum (DE) ; Ludwig-Maximilian-Universitaet Muenchen (DE) ; PD Dr. Dr. (DE) ; Schoch, Claudia (DE) ; Kern, Wolfgang (DE) Location/Qualifiers				
	1..1172				
	/organism="Homo sapiens"				



```

Db      1 CTTTCCGGAGACTGAGTGAAGGCCGAGATTCTTAAAGCCAGTGTAGAGACTA 60
Qy      61 TGTAGGCAAGAGTACTATAGAACCCCGAGAGGGTGAAGAGCAGCTGCCATATG 120
Db      61 TGTAGGCAAGAGTACTATAGAACCCCGAGAGGGTGAAGAGCAGCTGCCATATG 120
Qy      121 GCAGGTGAAGAAATTAATGAAGACTATCCAGTAAATTCACGAGTATTTCTGCGCTT 180
Db      121 GCAGGTGAAGAAATTAATGAAGACTATCCAGTAAATTCACGAGTATTTCTGCGCTT 180
Qy      181 GGAATTCATGTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db      181 GGAATTCATGTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Qy      241 GAGTGTGACAGAACTGATCCATCTTGAACAAGAAAGGATTTGTTCTGCAATAC 300
Db      241 GAGTGTGACAGAACTGATCCATCTTGAACAAGAAAGGATTTGTTCTGCAATAC 300
Qy      301 ACATTAATTCATGTTTGGGTTTATTTGGCAACCCAGAGATTATCCTTAAGAACAT 360
Db      301 ACATTAATTCATGTTTGGGTTTATTTGGCAACCCAGAGATTATCCTTAAGAACAT 360
Qy      361 CCAAGTAAACAGAAATGGAAGATGAGATATATATGAACAGAGTCAAGAAATTAACA 420
Db      361 CCAAGTAAACAGAAATGGAAGATGAGATATATATGAACAGAGTCAAGAAATTAACA 420
Qy      421 GACAAGAAAAAGGCTGCAAGCTGACAGAGTGCAGCTTCAAGATTTGTAATAAATGCC 480
Db      421 GACAAGAAAAAGGCTGCAAGCTGACAGAGTGCAGCTTCAAGATTTGTAATAAATGCC 480
Qy      481 CTCTGGAGAACCAAAATGGAAGATGCAAAAGTGCCTTAATTAAGAAAGTAAAGT 540
Db      481 CTCTGGAGAACCAAAATGGAAGATGCAAAAGTGCCTTAATTAAGAAAGTAAAGT 540
Qy      541 TAACCTTTGGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db      541 TAACCTTTGGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Qy      601 TATGTAAGCTAAATCTCTCTCCAAAGATCATTTATTTATGATTAAGACTGAGA 660
Db      601 TATGTAAGCTAAATCTCTCTCCAAAGATCATTTATTTATGATTAAGACTGAGA 660
Qy      661 TTTTAACATTTGATATATATATATATATATATATATATATATATATATATATAT 720
Db      661 TTTTAACATTTGATATATATATATATATATATATATATATATATATATATATAT 720
Qy      721 TCTTATATAGTCACTGCTGCAAGTACCATTTTAAAGCAGCTGAAATTTAAGTAA 780
Db      721 TCTTATATAGTCACTGCTGCAAGTACCATTTTAAAGCAGCTGAAATTTAAGTAA 780
Qy      781 ATGTTCTTTGTAACATTTGTAATTTTAAATGAATGAATGAATGAATGAATGAAT 840
Db      781 ATGTTCTTTGTAACATTTGTAATTTTAAATGAATGAATGAATGAATGAATGAAT 840
Qy      841 CTGAGGCTGAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 899
Db      841 CTGAGGCTGAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 899
Qy      900 TGACTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 959
Db      900 TGACTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 959
Qy      960 ATTACTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1019
Db      960 ATTACTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1019
Qy      1020 TTGGCAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1079
Db      1020 TTGGCAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1079
Qy      1080 GAAATTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1139
Db      1080 GAAATTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1139
Qy      1081 GAAATTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140

```

```

Qy      1140 CTGAAATATGTTGTAAAA 1156
Db      1141 CTGAAATATGTTGTAAAA 1157

RESULT 7
BC009589
LOCUS
DEFINITION
BC009589      1186 bp      mRNA      linear      PRI 29-JUN-2004
ACCESSION
BC009589
VERSION
BC009589.1 GI:16307025
KEYWORDS
MGC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 1186)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Datchenko, L., Martinis, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Sapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Umedin, T.B., Toshiyuki, S.,
Abramski, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Morley, K.C., Hale, S., Garcia, A.M., Gay, U.J., Hui, S.W.,
Vallalou, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Rodriguez, S.,
Boutard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalski, U., Smallos, D.E.,
Scherer, A., Schein, J.E., Jones, S.J., and Maitra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1186)
Strausberg, R.
Direct Submission
Submitted (29-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcabbs-remail.nih.gov
Tissue Procurement: ARCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAL Plate: 21 Row: b Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 27894372.
Location/Qualifiers
1. 1186
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:14659 IMAGE:4102498"
/tissue_type="Bone marrow, Chronic myelogenous leukemia"
FEATURES
source

```



## ORIGIN

91 GGAGGGTGAAGACAGAGCTGGCCATATGGCAGGTGAATAAATTTATGAAGACTATCCA 150

QY	151	GTAAATAATTCACAGTATTGTGCAGCGTTTGAGAAATTCACATTGGTGCCTGTGATACAGATG	210
Db	83	GTAAATAATTCACAGTATTTGTGCAGCGTTTGAGAAATTCACATTGGTGCCTGTGATACAGATG	142
QY	211	CTGAAGACCATGATGTCTGTGTTCTAGAAATGAGTGTGGCAGAAAGTTGGATCCACTTGA	270
Db	143	CTGAAGACCATGATGTCTGTGTTCTAGAAATGAGTGTGGCAGAAAGTTGGATCCACTTGA	202
QY	271	CAACCAAAAGTGGATNTGGTTTCTGCATATACATTTAAATTCAAATGTTTGGTTTATTTG	330
Db	203	CAACCAAAAGTGGATNTGGTTTCTGCATATACATTTAAATTCAAATGTTTGGTTTATTTG	262
QY	331	GCAACCCCAAGAGTTTAATCCTTAAGGAAATCATCAATGAAAACAGAAATTTGAAAAGATTCAGA	390
Db	263	GCAACCCCAAGAGTTTAATCCTTAAGGAAATCATCAATGAAAACAGAAATTTGAAAAGATTCAGA	322
QY	391	GTATATATGAACAGAGTCAAGGAAATTAACAGCAAGAAAAGGCTGGCAAGCTGGACAGA	450
Db	323	GTATATATGAACAGAGTCAAGGAAATTAACAGCAAGAAAAGGCTGGCAAGCTGGACAGA	382
QY	451	GGTCAGCTTCAAGATTTGTAAAAAATGCCCTCTGGGAAACCAAAATCGAAAAATGCATCA	510
Db	383	GGTCAGCTTCAAGATTTGTAAAAAATGCCCTCTGGGAAACCAAAATCGAAAAATGCATCA	442
QY	511	AAAGTTGCCAATTAAGGAAAAAGTAAAGTTAACTTTTGGTTTGATGTACATATTC	570
Db	443	AAAGTTGCCAATTAAGGAAAAAGTAAAGTTAACTTTTGGTTTGATGTACATATTC	502
QY	571	AAAAAGTAC-----	579
Db	503	AAAAAGTACATCTTCCCCCCCCCCCCCGCAAAATTAATTCGTGGCAGGGCAAGGTT	562
QY	580	-----ATTATATATGTATACAG-----TAATATGTAAAGTAAATAC-TTCC	621
Db	563	TAATATGTATCTTATTTATATATATGTAAATTCACAGTAAATATGTAAAGTAAATATCTTCC	622
QY	622	TCTGCAAAAGATCATTAATCTTATATGATTAGACATGAGAGATTTTAAATCATTTGATATATTA	682
Db	623	TCTGCAAAAGATCATTAATCTTATATGATTAGACATGAGAGATTTTAAATCATTTGATATATTA	682
QY	682	TATATTTATATATTTAACCATCTCTTGATGAGACTCTTATTTCTTTATATATAGTCAAGCTTG	741
Db	683	TATATTTATATATTTAACCATCTCTTGATGAGACTCTTATTTCTTTATATATAGTCAAGCTTG	742

**РЕСПУБЛИКА**

## REFERENCE

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT On Sep 16, 2003 this sequence version replaced gi:16740844

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Sep 16, 2003 this sequence version replaced

Contact: MGC help desk  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
BC Cancer Agency, Vancouver, BC, Canada  
info@cgsc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,  
Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth  
Featherstone, Malachi Griffith, Odi Griffith, Ran Guin, Nancy Liao,  
Teika Olson, Diana Palmquist, Mike R. Mayo, Josh Moran, Ryan Morin,  
Parvaneh Saeedi, JR Santos, Angela Petrescu, Anna Liisa Prabh, Duane Smalins, Jeff Stott, Mirzanda Trai, George Yang, Jacque Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAX Plate: 15 Row: a Column: 2  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.  
Location/Qualifiers

## FEATURES

source

1. 1168  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:9308 IMAGE:3906517"  
/tissue\_type="uterus, leiomyosarcoma"  
/clone\_id="NIH MGC\_71"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPOrt6"  
1. 1168  
/gene="CID"  
/note="Synonyms: MGC12261, MGC14659, SUNCOR"  
/db\_xref="LOCUSID:10438"  
/db\_xref="MIM:606997"  
49. 474  
/gene="CID"  
/product="nuclear DNA-binding protein"  
/protein\_id="AAH16284.1"  
/db\_xref="GI:16740845"  
/db\_xref="LOCUSID:10438"  
/db\_xref="MIM:606997"  
/translation="MAGEINBDYDVEIHEYLSPENSIGAVDEMLKTMMSVSRNEL  
OKLDPLEOKVDLVSATLNSFMWYLATGVNPKHPVQOEIRIRVYNNRVKEITD  
KKKAGKIDRGASRFRVKAALWEPKSKVANKSKSKS"

## gene

## CDS

## ORIGIN

## Query Match

Best Local Similarity 93.0%; Score 926; DB 9; Length 1168;  
Matches 1040; Conservative 0; Mismatches 10; Indels 68; Gaps 4;

QY 107 AGCTGCCATATATGTCAGTGAAGAAATTAATGAAGCTATCCAGTGAATTAATCAAGAT 166  
DB 38 AGTCAGCATATATGTCAGTGAAGAAATTAATGAAGCTATCCAGTGAATTAATCAAGAT 166  
QY 167 ATTGTCTAGCGCTTTGAAGATTCATGCTGCTGATGATGATGATGATGATGATGATGAT 226  
DB 98 ATTGTCTAGCGCTTTGAAGATTCATGCTGCTGATGATGATGATGATGATGATGATGAT 157  
QY 227 CTGTTTCTAGAAATGATGTTGTCAGAAAGTTGATCACTTGAACAAGCAAAAGTGAAT 286  
DB 158 CTGTTTCTAGAAATGATGTTGTCAGAAAGTTGATCACTTGAACAAGCAAAAGTGAAT 286  
QY 287 TGGTTTCTGATACACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 346  
DB 218 TGGTTTCTGATACACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 277  
QY 347 ATCTAAGAAATCCAGTAAACAGAAATGGAAGATCAAGATATATATGAACAGAG 406  
|||||

DB 278 ATCTAAGAAATCCAGTAAACAGAAATGGAAGATCAAGATATATATGAACAGAG 337  
QY 407 TCAGAAATTAACAGACAAAGAAAGCTGGCAAGCTGACAGAGTGCAGCTTCAAGAT 466  
DB 338 TCAGAAATTAACAGACAAAGAAAGCTGGCAAGCTGACAGAGTGCAGCTTCAAGAT 397  
QY 467 TTTGTAATAATGCTCTTGGGACCAAAATCGAAATAATGATCAAAAGTTCGCAATTAAG 526  
DB 398 TTTGTAATAATGCTCTTGGGACCAAAATCGAAATAATGATCAAAAGTTCGCAATTAAG 457  
QY 527 GAAAAAGTAAAGTTAACTTTTGGTTTGAATGACATATTCAAAAGTAC----- 579  
DB 458 GAAAAAGTAAAGTTAACTTTTGGTTTGAATGACATATTCAAAAGTACCTTCC 517  
QY 580 -----ATTA 583  
DB 518 CCCCCCCCCCGCAAAATATATCTGTGGCAGGCAAGCTTAAATGTTCTTTATTA 577  
QY 584 ATATGTAATCACA---GTATATGTAAGCTAATTC-TTCTCTCCAAAGATCATATTC 639  
DB 578 ATATGTAATTCACAGTAATATGTAAGCTAATTCCTCTCCAAAGATCATATTC 637  
QY 640 TTTATGTAATGACAGTGAAGATTTTAACTGTGATATATATATATATATATATATCA 699  
DB 638 TTTATGTAATGACAGTGAAGATTTTAACTGTGATATATATATATATATATATATCA 697  
QY 700 TCTCTGATGAGACCTTATATCTTTATATATATATATATATATATATATATATATAT 759  
DB 698 TCTCTGATGAGACCTTATATCTTTATATATATATATATATATATATATATATATAT 757  
QY 760 CAGCTGTAATTAAGTAAATGTTCTTGTAAACATTTGATCTATTTTAAATGAATA 819  
DB 758 CAGCTGTAATTAAGTAAATGTTCTTGTAAACATTTGATCTATTTTAAATGAATA 817  
QY 820 TCACCTTAAGATGATGATCTATCTGATGATGATGATGATGATGATGATGATGATGAT 879  
DB 818 TCACCTTAAGATGATGATCTATCTGATGATGATGATGATGATGATGATGATGATGAT 877  
QY 880 TCATATTAAGAAAGGTG-AAATGACTTAAATGTTCTTGTAAATGATGATGATGAT 938  
DB 878 TCATATTAAGAAAGGTGAAATGATGATGATGATGATGATGATGATGATGATGAT 937  
QY 939 CATGTTTCAATGATTTTGAAGATTAATGCTGTTGTTGATGATGATGATGATGATGAT 998  
DB 938 CATGTTTCAATGATTTTGAAGATTAATGCTGTTGTTGATGATGATGATGATGATGAT 997  
QY 999 TTTATGTTGATCTTAATTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1058  
DB 998 TTTATGTTGATCTTAATTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1057  
QY 1059 ATTCTCATTAATCTTGAATGAAAGTGAAGATGATGATGATGATGATGATGATGATGAT 1118  
DB 1058 ATTCTCATTAATCTTGAATGAAAGTGAAGATGATGATGATGATGATGATGATGATGAT 1117  
QY 1119 TGAATGATCTGCAATTAAGGCTGAAGATGTTGTAA 1156  
DB 1118 TGAATGATCTGCAATTAAGGCTGAAGATGTTGTAA 1155

## RESULT 9

BC009584

LOCUS

## DEFINITION

Homo sapiens nuclear DNA-binding protein, transcript variant 1, BC009584

## ACCESSION

Homo sapiens nuclear DNA-binding protein, transcript variant 1, BC009584

## VERSION

BC009584.1 GI:16307017

## KEYWORDS

MGC.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1195)

## AUTHORS

Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G.,



TITLE	JOURNAL	PIUMED	REFERENCE	AUTHORS	JOURNAL	REMARK	COMMENT
Clasauer, R.D., Collins, F.S., Wagner, L., Shemer, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stopleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scherch, T.E., Brownstein, M.J., Ueding, T.B., Toshitsuki, S., Carinci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosek, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worcester, K.C., Hale, S., Garcia, A.M., Gay, L.J., Halik, S.W., Viall, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Patton, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Bakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, R., Myers, R.M., Butlerfield, Y.S., Krzywnicki, M.I., Skalka, U., Smalins, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.	Submitted (29-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a>	Contact: MGC help desk	Email: <a href="mailto:cgagphs-remail.nih.gov">cgagphs-remail.nih.gov</a>	Tissue Procurement: ATCC	CNA Library Preparation: CLONTECH Laboratories, Inc.	CNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305	Web site: <a href="http://www.shgc.stanford.edu">http://www.shgc.stanford.edu</a>	Contact: (Dickson, Mark) <a href="mailto:mcdpaxil@stanford.edu">mcdpaxil@stanford.edu</a>	Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a>	Series: IRAL Plate: 21 Row, n Column: 3.	Location/Qualifiers	1. .1195
/organism="Homo sapiens"	/mol_type="mRNA"	/db_xref="taxon:9606"	/clone="MGC:14647 IMAGE:4096552"	/tissue_type="Bone marrow, chronic myelogenous leukemia"	/clone_lib="NIH MGC_54"	/lab_host="DH10B"	/note="vector: pDNR-LIB"
1. .1195	/gene="CID"	/note="synonyms: MGCI2261, MGCI4659, SUNCOR"	/db_xref="LOCUSID:10438"	/db_xref="MIM:606997"	56. .481	/gene="CID"	/codon_start=1
/product="nuclear DNA-binding protein"	/protein_id="AAH09584.1"	/db_xref="GI:16307018"	/db_xref="LOCUSID:10438"	/db_xref="MIM:606997"	/translation="IMAGEINDBEDPVEIHETLSAPENSIGAVDEMLKTMNSVSHRELLQKDLPEQAKVDLSATLNSMFVYATLQGVNKEHPVQCEIRIRIVYNNRVEITDDKKAAGKIDRGAASRFVKNALWEPKSKNASKVANKGSKS"	79.84; Score 922.4; DB 9; Length 1195;	

	Beef	Local	Similarity	92.8%	Pred. No. 3.6e-155;	Matches 1039;	Conservative 0;	Mismatches 11;	Indels 70;	Gaps 4
QY	107	AGCTGCCATATATGCGAGGTGAGAAATTTAATGAAAGCTATCCAGTAGAAATTCAGAGT	166							
Db	45	AGTCAGCCATATATGCGAGGTGAGAAATTTAATGAAAGCTATCCAGTAGAAATTCAGAGT	104							
QY	167	ATTGTGACGCGTTTGAAGAAATTCATTTGGTGTCTGTGTGATGAGATGCTGAAAGCATGATGT	226							
Db	105	ATTGTGACGCGTTTGAAGAAATTCATTTGGTGTCTGTGTGATGAGATGCTGAAAGCATGATGT	164							
QY	227	CTGTTTCTAGAAATGATGTTGTGCAAGATTGGATCCACTTGAAACAGCAAAAGTGAT	286							
Db	165	CTGTTTCTAGAAATGATGTTGTGCAAGATTGGATCCACTTGAAACAGCAAAAGTGAT	224							
QY	287	TGGTTTCTGCAATACATTTAAATTCATTTGGTTTATTTGGCAACCCAAAGAGTTA	346							
Db	225	TGGTTTCTGCAATACATTTAAATTCATTTGGTTTATTTGGCAACCCAAAGAGTTA	284							
QY	347	ATCTTAAGGAACATCCAGTAAACAGAAATTTGGAAAGATCAGATATATGAAACAGAG	406							
Db	285	ATCTTAAGGAACATCCAGTAAACAGAAATTTGGAAAGATCAGATATATGAAACAGAG	344							
QY	407	TCAGGAATATATACAGACAGAAAGAGGTGGCAAGCTGACAGAGCTCAGCTTCAAGAT	466							
Db	345	TCAGGAATATATACAGACAGAAAGAGGTGGCAAGCTGACAGAGCTCAGCTTCAAGAT	404							
QY	467	TTGTAAATAATATGCGCTCTGGGAAACCAAAATGCAATTCATAAAATTTGGCAATTAAG	526							
Db	405	TTGTAAATAATATGCGCTCTGGGAAACCAAAATGCAATTCATAAAATTTGGCAATTAAG	464							
QY	527	GAATAAGTAAGTTAACTTTTGGTGTGTGATGATACATATATTCAAAAGTAC-----	579							
Db	465	GAATAAGTAAGTTAACTTTTGGTGTGTGATGATACATATATTCAAAAGTACCTTCC	524							
QY	580	-----AT 581								
Db	525	CCCCCCCCCCCCCGAAATTAATTCGTGCAAGGCGAAGTTTAAATGTGTCTTAT	584							
QY	582	TATATGTAATCACA---GTATATATGTAAAGCTAAATPAC-TTCCTCTCCAAAGATCATTA	637							
Db	585	TATATGTAATCACAAGTAAATATGTAAAGCTAAATCTTCTCTCCAAAGATCATTA	644							
QY	638	TCTTATATGATAGCACTGAGAGTTTAACTTGTGATATAATTAATTTAAATTTAC	697							
Db	645	TCTTATATGATAGCACTGAGAGTTTAACTTGTGATATAATTAATTTAAATTTAC	704							
QY	698	CATCTCTGATGAGACCTCTTAATTTCTTTATATAGGTCACTCTTGCAAGTACCATTTTATA	757							
Db	705	CATCTCTGATGAGACCTCTTAATTTCTTTATATAGGTCACTCTTGCAAGTACCATTTTATA	764							
QY	758	AGCAGCTGTGAAATTTAAGTAAATGTCTTTGAAACATTTGTAATTTTAAATGAT	817							
Db	765	AGCAGCTGTGAAATTTAAGTAAATGTCTTTGAAACATTTGTAATTTTAAATGAT	824							
QY	818	AATGACCTTATGAGATATGCTATCTGTAAGCTGAAATTAATAGTACATCTGTTTCACTA	877							
Db	825	AATGACCTTATGAGATATGCTATCTGTAAGCTGAAATTAATAGTACATCTGTTTCACTA	884							
QY	878	TATGATATTAAGAAAGCGTG-AATGACTTAAATGTCAATTTTTCTGTATAGATACTTT	936							
Db	885	TATGATATTAAGAAAGCGTGAAATGACTTAAATGTCAATTTTTCTGTATAGATACTTT	944							
QY	937	ATCATGTTTTCACATTTTGAAGATTTACGTGTTGTTGATATTCMAAGTGTGAACCTTAA	996							
Db	945	ATCATGTTTTCATATTTTGAAGATTTACGTGTTGTTGATATTCMAAGTGTGAACCTTAA	1004							
QY	997	AGTTATATGTTGATCTTAAATCTTGGCAAGTTCCTCTATATGTCCCAATTTAAATTTAAAT	1056							
Db	1005	ACTTATATGTTGATCTTAAATCTTGGCAAGTTCCTCTATATGTCCCAATTTAAATTTAAAT	1064							
QY	1057	ACATTCTCATTAACCTTATAGTGGAAATTAAGTTGATATGTTGATGATGAATTTTGGCAT	1116							

Db 1065 ACATTCTCATTAAGTGAAGGTAAGTGTATGTATGATGATGAATTTGGCAT 1124  
QY 1117 GATGCTGACTCTCAATTAAGGCTGAAGATGTTGTAAA 1156  
Db 1125 GATGACTGACTCTCAATTAAGGCTGAAGATGTTGTATTA 1164

RESULT 10  
BC005235  
LOCUS  
DEFINITION Homo sapiens nuclear DNA-binding protein, transcript variant 1,  
BC005235  
ACCESSION BC005235.1 GI:13528872  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1199)  
Strausberg, R.D., Collins, F.S., Wagner, L.H., Derge, J.G.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Datchenko, L., Matusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Shiepel, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schaefer, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Abrams, R.D., Mullany, S.J., Bosak, S.A., McMan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richard, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulik, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Sanchez, A., Whiting, M., Madan, A., Madan, A., Rodriguez, S.,  
Bouffard, G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smalins, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
human and mouse cDNA sequences  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 1199)  
Strausberg, R.  
Direct Submission  
Submitted (27-MAR-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
DNA Sequencing By: The I.M.A.G.E. Consortium (LMLU)  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcg@paxil.stanford.edu](mailto:mcg@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

FEATURES  
SOURCE  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LMLU at: <http://image.lml.gov>  
Series: IRAL Plate: 16 Row: 1 Column: 7  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 27894371.  
Location/Qualifiers  
1..1199  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:12261 IMAGE:3930648"

gene  
/tissue\_type="Lung, mucocypidermoid carcinoma"  
/clone\_lib="NIH MGC\_59"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
1..1199  
/gene="C1D"  
/note="synonyms: MGCL12261, MGCL14659, SUNCOR"  
/db\_xref="locusid:10438"  
/db\_xref="MIM:60697"  
64..469  
/gene="C1D"  
/codon\_start=1  
/product="nuclear DNA-binding protein"  
/protein\_id="AAH05235.1"  
/db\_xref="GI:13528873"  
/db\_xref="LOC810:10438"  
/db\_xref="MIM:60697"  
/translation="MAGEB1NEDYPIVEIHEYLAFENSIGAVDEMKTMMVSRLNELL  
OKLDPLEQAKVDIVSAVYTLNMFVYATQVNPKEHPVKQELERIRVYNNRVKEITD  
KKKAGKLDKRGASRPVKNALMEPKSNKNAKVKKSKS"

CDS  
Query Match 79.6%; Score 919.8; DB 9; Length 1199;  
Best Local Similarity 92.6%; Pred. No. 1e-154;  
Matches 1038; Conservative 0; Mismatches 12; Indels 71; Gaps 4;

ORIGIN  
107 AGCTGGCCATATAGCAGGTGAAGAAATTAATGAACTATCCATGAAATTCACGAGT 166  
53 AGTCAGCATTAATGAGGAGTGAAGAAATTAATGAACTATCCATGAAATTCACGAGT 112  
167 ATTGTCAGGCTTGAATTCATTCATTCGCTGCTGATGATGATGCTGAAGACCATGAT 226  
113 ATTGTCAGGCTTGAATTCATTCATTCGCTGCTGATGATGATGCTGAAGACCATGAT 172  
227 CTGTTCTGAATATGATGTTGTCAGAAAGTGGATCCACTGAACCAACCAAGTGGATT 286  
173 CTGTTCTGAATATGATGTTGTCAGAAAGTGGATCCACTGAACCAACCAAGTGGATT 232  
287 TGGTTCTGATACATTAATTCATGTTTGGGTTTATTTGGCAACCAAGTGGATT 346  
233 TGGTTCTGATACATTAATTCATGTTTGGGTTTATTTGGCAACCAAGTGGATT 292  
347 ATCTTAAGGAACATCCAGTAAACAGAAATTTGAAAGATGATATTTGAACAGAG 406  
293 ATCTTAAGGAACATCCAGTAAACAGAAATTTGAAAGATGATATTTGAACAGAG 352  
407 TCAAGAAATTAACAGAAAGAAAGCTGCAAGCTGACAGAGTGCAGCTTCAAGAT 466  
353 TCAAGAAATTAACAGAAAGAAAGCTGCAAGCTGACAGAGTGCAGCTTCAAGAT 412  
467 TTGTAAATAAGCCCTCTGGGAACCAAAATGCAAAATGCAAAAGTTGCCAATTAAG 526  
413 TTGTAAATAAGCCCTCTGGGAACCAAAATGCAAAATGCAAAAGTTGCCAATTAAG 472  
527 GAAAAAGTAAAGTTAACTTTTGGTTTGAATGATACATATCAAAAAGTAC----- 579  
473 GAAAAAGTAAAGTTAACTTTTGGTTTGAATGATACATATCAAAAAGTACCTTCC 532  
580 -----A 580  
533 CCCCCCCCCCCCCGCAAAATTAATTCGTGCGAGGCAAGTTAAATGTTCTTA 592  
581 TTAATATGATACACA---GTAATATGTAAGCTAAATAC-TTCCTCCAAAGATCAT 636  
593 TTAATATGTAATTCACAGTAATTAATGTAAGCTAAATCTTCTCCCAAGATCAT 652  
637 ATCTTAATGATAGCACTGAGATTTTAACATTTGATATTAATTAATTAATTA 696  
653 ATCTTAATGATAGCACTGAGATTTTAACATTTGATATTAATTAATTAATTA 712  
697 CCATCTCTTGAAGAGCTTTATTTCTTAATTAATGATGCTTCCAGATCATTTAT 756  
713 CCATCTCTTGAAGAGCTTTATTTCTTAATTAATGATGCTTCCAGATCATTTAT 772



Em:AA442782 Em:AA305785 Em:AI804147 Em:AI809071  
Em:AI800846 Em:AI351827 Em:AA805413 Em:AA835937 Em:NA0848  
Em:AI553826 Em:AA458588 Em:AI881757 Em:AA42131 Em:AI335349  
Em:AI1224 Em:AW118590 Em:AA169841 Em:AA431129 Em:AI355430  
Em:AI243111 Em:AA424253 Em:AM02457 Em:AI699498  
Em:AI989469 Em:AI310127 Em:AA118433 Em:AI123353  
Em:AI800532 Em:AI350721 Em:AA460755 Em:R68158 Em:AA861428  
Em:AI273315 Em:AA424184 Em:AI375521 Em:AI027628  
Em:AI301876 Em:AA476674 Em:AI816911 Em:AI476631  
Em:AI476155 Em:AI435863 Em:AI458192 Em:AI04208  
Em:AA340622 Em:AI379715 Em:AI671948 Em:W01341 Em:AM992440  
/evidence=not experimental  
/complement(54351)  
/gene="DA19501.1"  
complement(54383). 54386)  
/gene="DA19501.1"  
54386. 54689  
/note="match: STS: Em:G23952"  
complement(54999. 55406)  
/gene="DA19501.1"  
/note="match: GSS: Em:AQ461100"  
complement(55051. 55386)  
/gene="DA19501.1"  
/note="possible pseudogene  
match: proteins: Tr:Q13901"  
/codon start=1  
/evidence=not experimental  
/product="ba19501.1 (novel protein similar to CID  
DNA-binding protein)"  
/protein\_id="CAD13204.1"  
/db\_xref="GI:17384063"  
/translation="MLKTSVSRNLEIÖKDLPELQARVDIVSAVTLSNMTVTLAQ  
GVNPKPVPQOELEIRIVYMNRYKETITDKKRGKLDRAARAFVNAALMEPKRNASK  
VANKGSKS"  
55412. 55487  
/note="match: GSS: Em:AQ63761 Em:AQ050503"  
61512. 61611  
/note="50 copies 2 mer aa 61% conserved"  
68837. 68982  
/note="LIPAs repeat: matches 5997. 6143 of consensus"  
71617. 71998  
/note="LIPB3 repeat: matches 5769. 6150 of consensus"  
72363. 72506  
/note="72 copies 2 mer ta 77% conserved"  
78102. 78408  
/note="match: GSS: Em:AQ628574"  
complement(78524. 78950)  
/note="match: GSS: Em:AQ401519"  
78659. 78758  
/note="LIMB3 repeat: matches 102. 199 of consensus"  
79868. 80242  
/note="LTR38 repeat: matches 129. 503 of consensus"  
80250. 84310  
/note="LIPAs repeat: matches 2098. 6155 of consensus"  
84306. 84837  
/note="LIPAs repeat: matches 1407. 1939 of consensus"  
join(87084. 87121.87983. 88037.109294. 109356,  
111406. 111627.115595. 115629.118165. 118302,  
126770. 127043.127968. 128040.177990. 178100)  
/gene="BA479G22.1"  
join(87084. 87121.87983. 88037.109294. 109356,  
126770. 111627.115595. 115629.118165. 118302,  
126770. 127043.127968. 128040.177990. 178100)  
/gene="BA479G22.1"  
/product="DA19501.3 (novel protein)"  
/note="isoform 2  
continues as BA453G10.1 in Em:ALJ56053  
match: CDNAS: Em:AK023093  
match: ESTs: Em:AA421130 Em:AA609088"  
/evidence=not experimental  
join(87084. 87121.87983. 88037.109294. 109356,  
111406. 111627.126770. 127043.127968. 128040. 177990. 178100)  
177990. 178100)

[illegible]

Qy 858 AGGACATCTGTTTCTACTATATGATATTAAGAAAGGCTG-AATGACTTAATGTCATT 916  
 Db 54663 AGGCACATCTGTTTCTACTATATGATATTAAGAAAGGCTGAAAGACTTAATGTCATT 54604  
 Qy 917 TTTTCTGATAGATCTTATATCATGTTTTCATGATTTTAGAATTAATCTGCTTGTGAT 976  
 Db 54603 TTTTCTGATAGATCTTATATCATGTTTTCATGATTTTAGAATTAATCTGCTTGTGAT 54544  
 Qy 977 ATTCAAGTGTAAAGTAAAGTTATGTTGTTGTTTAACTTTCCTGCGATGTTGCTCTA 1036  
 Db 54543 ATTCAAGTGTAAAGTAAAGTTATGTTGTTGTTTAACTTTCCTGCGATGTTGCTCTA 54484  
 Qy 1037 TGTCCCATTTAAATTAATATATCATCTTCATTAATTAGATGGGAATTAAGTTGTATGT 1096  
 Db 54483 TGTCCCATTTAAATTAATATATCATCTTCATTAATTAGATGGGAATTAAGTTGTATGT 54424  
 Qy 1097 TGATGATGTAATTTTGCGATGATGATCTGTACTCTCAATAAAGCTGAAATGTTGTAATA 1156  
 Db 54423 TGATGATGTAATTTTGCGATGATGATCTGTACTCTCAATAAAGCTGAAATGTTGTAATA 54364

RESULT 12  
 AC068741/c 195426 bp DNA linear HTG 22-NOV-2000  
 LOCUS Homo sapiens clone RP11-777E12, WORKING DRAFT SEQUENCE, 8 unordered  
 DEFINITION pieces.  
 AC068741 GI:11276256  
 AC068741.3 HTG: HTGS\_PHSB1; HTGS\_DRAFT.  
 VERSION HTG: HTGS\_PHSB1; HTGS\_DRAFT.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 195426)  
 AUTHORS Birren, B., Linton, L., Nuebaum, C. and Lander, E.  
 TITLE Homo sapiens, clone RP11-777E12  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 195426)  
 AUTHORS Birren, B., Linton, L., Nuebaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Baschien, V., Beda, F.,  
 Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karas, A.,  
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehotzky, J.,  
 Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, J.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R.,  
 Melidim, J., Meneus, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Teste, S., Theodore, J., Tittell, A., Travers, M., Triggillo, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 DIRECT SUBMISSION

TITLE JOURNAL  
 COMMENT Submitted (08-MAY-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Nov 22, 2000 this sequence version replaced g1:9369524.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence submissions@genome.wi.mit.edu  
 ----- Project Information

Center project name: 18655  
 Center clone name: 777 E.12  
 ----- Summary Statistics  
 Sequencing vector: M13; M7815; 30% of reads  
 Sequencing vector: Plasmid; n/a; 70% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 191262 bases at least Q40  
 Consensus quality: 193601 bases at least Q30  
 Consensus quality: 194390 bases at least Q20  
 Insert size: 182000; agarose-fp  
 Insert size: 194726; sum-of-ctnigs  
 Quality coverage: 6.9 in Q20 bases, agarose-fp  
 Quality coverage: 6.5 in Q20 b.  
 NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 731: contig of 731 bp in length  
 \* 732 831: gap of 100 bp  
 \* 832 3248: contig of 2417 bp in length  
 \* 3249 3348: gap of 100 bp  
 \* 3349 8620: contig of 5272 bp in length  
 \* 8621 8720: gap of 100 bp  
 \* 8721 17691: contig of 8971 bp in length  
 \* 17692 17791: gap of 100 bp  
 \* 17792 25774: contig of 7983 bp in length  
 \* 25775 25874: gap of 100 bp  
 \* 25875 66476: contig of 40602 bp in length  
 \* 66477 66576: gap of 100 bp  
 \* 66577 128195: contig of 61619 bp in length  
 \* 128196 128295: gap of 100 bp  
 \* 128296 195426: contig of 67131 bp in length.  
 Location/Qualifiers  
 1. 195426  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="RP11-777E12"  
 /clone\_lib="RP11 Human Male BAC"  
 1. 731  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"

misc\_feature  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"  
 misc\_feature  
 832..3248  
 /note="assembly\_fragment"  
 misc\_feature  
 3349..8620  
 /note="assembly\_fragment"  
 misc\_feature  
 8721..17691  
 /note="assembly\_fragment"  
 misc\_feature  
 17792..25774  
 /note="assembly\_fragment"  
 misc\_feature  
 25875..66476  
 /note="assembly\_fragment"  
 misc\_feature  
 66577..128195  
 /note="assembly\_fragment"  
 misc\_feature  
 128296..195426  
 /note="assembly\_fragment  
 clone\_end:T7  
 vector\_side:right"

## ORIGIN

Query Match 77.9%; Score 900.4; DB 2; Length 195426;  
 Best Local Similarity 91.2%; Pred. No. 1.1e-151;  
 Matches 1040; Conservative 0; Mismatches 26; Indels 74; Gaps 5;  
 Qy 91 GGAGGCTGAGGACAGACAGCTGCGCATATATGCGAGTGAAGAAATTAATGACATATCCA 150  
 Db 98376 GGATGGAAGGCGCTGATGACGCAATACGCGACGATGAAGAAATTAATGACATATCCA 98317

```

Qy 151 GTAGAAATTCAGAGATTTGTCACGCTTTGAGAAATTCATTGCTGTGATGATG 210
Db 98316 GTAGAAATTCAGAGATTTGTCACGCTTTGAGAAATTCATTGCTGTGATGATG 98257
Qy 211 CTGAAGACCATGATGCTGTTCTTGAATGATGTTGGCAGAGTTGGATCCCTTGA 270
Db 98256 CTGAAGACCATGATGCTGTTCTTGAATGATGTTGGCAGAGTTGGATCCCTTGA 98197
Qy 271 CAGCAAAAGTGGATTTGGTTTCTGCTATACATTAAATTCATGTTTGGTTTATTTG 330
Db 98196 CAGCAAAAGTGGATTTGGTTTCTGCTATACATTAAATTCATGTTTGGTTTATTTG 98137
Qy 331 GCAACCCAGAGATTAATCCTTAAGAAATCATCTAGTAAACAGAAATTTGAAATACAG 390
Db 98136 GCAACCCAGAGATTAATCCTTAAGAAATCATCTAGTAAACAGAAATTTGAAATACAG 98077
Qy 391 GTATATATGACACAGTCAGAGAAATACAGACAGAGAAAGGCTGGCAAGCTGACAG 450
Db 98076 GTATATATGACACAGTCAGAGAAATACAGACAGAGAAAGGCTGGCAAGCTGACAG 98017
Qy 451 GGTGACCTTCAGATTTGTAAAAAATGCTCTGGGAACCAAAATCGAAAAATGCATCA 510
Db 98016 GGTGACCTTCAGATTTGTAAAAAATGCTCTGGGAACCAAAATCGAAAAATGCATCA 97957
Qy 511 AAGTTGCCAATAAGAAAAAGTAACTTTTGGTTGATGTACATATTC 570
Db 97956 AAGTTGCCAATAAGAAAAAGTAACTTTTGGTTGATGTACATATTC 97897
Qy 571 AAAAGTAC----- 579
Db 97896 AAAAGTACATTTCCCGCCCCCCCCCCCCCAAGAAATTCCTGGCAGACAG 97837
Qy 580 -----ATTATATGTATACAG--TATATGTAAAGCTAAATAC--TT 619
Db 97836 TTTAAATGCTTTCTTATTAATATGTAATTCACAGTAAATATGTAAGCTAAATACCTTT 97777
Qy 620 CCTCTCCAAAGATCATTTCTTATTTGATACAGAGATTTTAACTTTGATATAT 679
Db 97776 CCTCTCCAAAGATCATTTCTTATTTGATACAGAGATTTTAACTTTGATATAT 97717
Qy 680 TATATTTTAAATTTTACCTTC--TTGATGAGCTTTATTTCTTAAATAGTCACT 737
Db 97716 TATATTTTAAATTTTACCTTCCTTTGATGAGCTTTATTTCTTAAATAGTCACT 97657
Qy 738 CTTCGAGTACATTTTAAAGCAGCTGTAATTTAAGTAAATGTTCTTTGTAACAT 797
Db 97656 CTTCGAGTACATTTTAAAGCAGCTGTAATTTAAGTAAATGTTCTTTGTAACAT 97597
Qy 798 TTGTACTATTTTAAATGAATATGACCTTATGAAGTATGCTATCTGAGGCTGAAATAT 857
Db 97596 TTGTACTATTTTAAATGAATATGACCTTATGAAGTATGCTATCTGAGGCTGAAATAT 97537
Qy 858 AGTACATCTGTTTCACTATATGATATTAAGAAAGGCTG--AATGACTTAAATGTTCAAT 916
Db 97536 AGGACATCTGTTTCACTATATGATATTAAGAAAGGCTG--AATGACTTAAATGTTCAAT 97477
Qy 917 TTTTTCGTATGATACCTTATATCATGTTTCAATGATTTAGAAATACCTTTGTTGAT 976
Db 97476 TTTTTCGTATGATACCTTATATCATGTTTCAATGATTTAGAAATACCTTTGTTGAT 97417
Qy 977 ATTCAAGTGTGAAACTTAAAGTTATGTTGTTACTTTAAATTTTGGCATGTTGCTCTTA 1036
Db 97416 ATTCAAGTGTGAAACTTAAAGTTATGTTGTTACTTTAAATTTTGGCATGTTGCTCTTA 97357
Qy 1037 TGTCTCCATTTAAATATATACATTCATTAATCTTAAAGTGGAAATAGAGTTGATGT 1096
Db 97356 TGTCTCCATTTAAATATATACATTCATTAATCTTAAAGTGGAAATAGAGTTGATGT 97297
Qy 1097 TGAATGATGAATTTTGGCATGATGCTGTAATCTCAATAAAGCTGAAATGTTGTAATA 1156
Db 97296 TGAATGATGAATTTTGGCATGATGCTGTAATCTCAATAAAGCTGAAATGTTGTAATA 97237

```

```

RESULT 13
AC079112/c
LOCUS
DEFINITION
AC079112
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155888)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
9847074
2 (bases 1 to 155888)
Goyea, E., Mishra, S., and Kozlowski, A.
The sequence of Homo sapiens BAC clone RP11-67K11
Unpublished
3 (bases 1 to 155888)
Waterston, R.H.
Direct Submission
Submitted (18-AUG-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 155888)
Waterston, R.H.
Direct Submission
Submitted (16-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 155888)
Waterston, R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 16, 2001 this sequence version replaced gi:13752170.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
Summary Statistics
Center project name: H_NH0067K11

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RP11-67K11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frenken, E., Tateo, M., Cacanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome

libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-30P5; the clone sequenced to the right is RP11-474G23. Actual start of this clone is at base position 1 of RP11-67K11; actual end is at base position 155888 of RP11-67K11.

FEATURES

source

1. 155888  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-67K11"  
/clone\_1kb="RPC1-11"  
328..526  
/rpt\_family="MER1\_type"  
538..645  
/rpt\_family="MIR"  
1745..2080  
/rpt\_family="ERV1"  
2082..2215  
/rpt\_family="Alu"  
2219..2389  
/rpt\_family="ERV1"  
2390..2821  
/rpt\_family="ERV1"  
3746..3978  
/rpt\_family="MIR"  
4181..4281  
/rpt\_family="MIR"  
4536..4849  
/rpt\_family="Alu"  
5603..5990  
/rpt\_family="MALR"  
5993..7584  
/rpt\_family="MALR"  
7585..7708  
/rpt\_family="Alu"  
7709..7824  
/rpt\_family="MALR"  
7825..7905  
/rpt\_family="MALR"  
7907..8614  
/rpt\_family="L1"  
8615..10087  
/rpt\_family="L1"  
10088..10108  
/rpt\_family="AT\_rich"  
10172..10432  
/rpt\_family="Alu"  
10433..10741  
/rpt\_family="MALR"  
11491..11646  
/rpt\_family="MIR"  
12515..12833  
/rpt\_family="Alu"  
12910..13181  
/rpt\_family="Alu"  
13182..13483  
/rpt\_family="Alu"  
13749..14156  
/rpt\_family="L2"  
14158..14654  
/rpt\_family="MALR"  
14763..14928  
/rpt\_family="MALR"  
14990..15229  
/rpt\_family="MALR"

repeat\_region 15940..17923  
/rpt\_family="L1"  
repeat\_region 18057..18200  
/rpt\_family="L1"  
repeat\_region 18806..18917  
/rpt\_family="GA-rich"  
repeat\_region 19295..19331  
/rpt\_family="AT\_rich"  
repeat\_region 20065..20084  
/rpt\_family="TC)n"  
repeat\_region 20091..20135  
/rpt\_family="TC)n"  
repeat\_region 20136..20179  
/rpt\_family="CT-rich"  
repeat\_region 21523..21686  
/rpt\_family="MER1\_type"  
repeat\_region 21829..22040  
/rpt\_family="MIR"  
repeat\_region 22370..22664  
/rpt\_family="L1"  
repeat\_region 22757..22858  
/rpt\_family="L1"  
repeat\_region 23117..23428  
/rpt\_family="Alu"  
repeat\_region 23565..24609  
/rpt\_family="L1"  
repeat\_region 25098..25143  
/rpt\_family="AT\_rich"  
repeat\_region 25479..25657  
/rpt\_family="L2"  
repeat\_region 25658..25929  
/rpt\_family="Alu"  
repeat\_region 25930..26003  
/rpt\_family="L2"  
repeat\_region 26003..26168  
/rpt\_family="MIR"  
repeat\_region 26390..26739  
/rpt\_family="MER2\_type"  
repeat\_region 26789..27175  
/rpt\_family="MALR"  
repeat\_region 27258..27303  
/rpt\_family="CA)n"  
repeat\_region 27536..27604  
/rpt\_family="MIR"  
repeat\_region 27589..27627  
/rpt\_family="L2"  
repeat\_region 28276..28369  
/rpt\_family="MIR"  
repeat\_region 28432..28836  
/rpt\_family="L2"  
repeat\_region 29950..30197  
/rpt\_family="MIR"  
repeat\_region 30283..30327  
/rpt\_family="CA)n"  
repeat\_region 30417..30565  
/rpt\_family="ERV1"  
repeat\_region 31120..31275

Query Match 56.7%; Score 655.8; DB 9; Length 155888;  
Best Local Similarity 90.7%; Pred. No. 6,4e-108;  
Matches 772; Conservative 0; Mismatches 7; Indels 72; Gaps 4;

QY 378 GGAAAGATCAGATATATATGACAGAGTCAGAAATTAACGACAGAGAAAAAGCTGG 437  
DB 116453 GGAAAGATCAGATATATATGACAGAGTCAGAAATTAACGACAGAGAAAAAGCTGG 116394  
QY 438 CAAGCTGACAGAGGTCAGAGCTTCAAGATTGTAATAAATCCCTCGGAACCAAAATC 497  
DB 116393 CAAGCTGACAGAGGTCAGAGCTTCAAGATTGTAATAAATCCCTCGGAACCAAAATC 116334  
QY 498 GAAATATGATCAAAAGTTGCCAATAAGAAAAAGTAAAGTTTCTTTGTTTGA 557  
DB 116333 GAAATATGATCAAAAGTTGCCAATAAGAAAAAGTAAAGTTTCTTTGTTTGA 116274



```

QY      558  TGTACACATATTCGAAAGTAC-----579
DB      116273 TGTACACATATTCGAAAGTACCTCCCGCCCCCGCCCCCGCAAAATATTCGTG 116214
QY      580  -----ATTAATATGTAATCACA---GTAATATCTAA 607
DB      116213 TGGCAGGCGCAAGTTAAATGTTCTTATTAATATGTAATTCACAGTAATATGTA 116154
QY      608  AGCTAAATATC-TCCTCTCCCAAGATCATTTCTTTATTTGATTTAGCATGAGATTTTAA 666
DB      116153 AGCTAAATATCTTCCTCCCAAGATCATTTCTTTATTTGATTTAGCATGAGATTTTAA 116094
QY      667  CATTTGATATATATATATATATATATATATATATATATATATATATATATATATATATAT 726
DB      116093 CATTTGATATATATATATATATATATATATATATATATATATATATATATATATATAT 116034
QY      727  TATAGTCAGTCTGCAAGTACCATTTTATTAAGAGCTGTGAATTTAAAGTGAATGTTTC 786
DB      116033 TATAGTCAGTCTGCAAGTACCATTTTATTAAGAGCTGTGAATTTAAAGTGAATGTTTC 115974
QY      787  TTTGTAAACATTTGTACTATTTTAAATGAATTAATGACCTTATTAAGATATGCTATCTGTA 846
DB      115973 TTTGTAAACATTTGTACTATTTTAAATGAATTAATGACCTTATTAAGATATGCTATCTGTA 115914
QY      847  GCTGAATTTATAGTACATCTGTTTTCATTAATGATATTAATTAAGAAAGCTGTG-AATGACTT 905
DB      115913 GCTGAATTTATAGTACATCTGTTTTCATTAATGATATTAATTAAGAAAGCTGTGAATGACTT 115854
QY      906  AAATGTTCATTTTCTGTATGATATCTTAATGATATTAATGATATTAATGATATTAATGAT 965
DB      115853 AAATGTTCATTTTCTGTATGATATCTTAATGATATTAATGATATTAATGATATTAATG 115794
QY      966  GCTTTGTATATTCGAAAGTGAACCTAAAGTTATGTTGATCTTAAATTTCTTGCA 1025
DB      115793 GCTTTGTATATTCGAAAGTGAACCTAAAGTTATGTTGATCTTAAATTTCTTGCA 115734
QY      1026  TGTTCCTCTATGTCCTCATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1085
DB      115733 TGTTCCTCTATGTCCTCATTTAAATTAATTAATTAATTAATTAATTAATTAATTA 115674
QY      1086  AGCTTGTATGTTGATGATGAATTTTGGCAGTACTGATCTGCAATTAAGGCTGAAA 1145
DB      115673 AGCTTGTATGTTGATGATGAATTTTGGCAGTACTGATCTGCAATTAAGGCTGAAA 115614
QY      1146  ATGTGTAAAA 1156
DB      115613 ATGTGTAAAA 115603

```

RESULT 14  
 AL356095/c  
 LOCUS  
 DEFINITION Human DNA sequence from clone Rpl1-369J21 on chromosome 10.  
 ACCESSION AL356095  
 VERSION AL356095.11 GI:12331029  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteleia; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 172079)  
 AUTHORS Brown, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 On Jan 22, 2001 this sequence version replaced gi:11863400.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the "unseq" feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10> Rpl1-369J21 is from the library RPl1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: PBACE3.6  
 This sequence is the entire insert of clone Rpl1-369J21 The true right end of clone Rpl1-479017 is at 65134 in this sequence.

FEATURES  
 source  
 Location/Qualifiers  
 1..172079  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="Rpl1-369J21"  
 /clone\_1fb="RPl1-11.2"  
 7..431  
 /note="match: GSS: Em:AQ191918"  
 25..413  
 /note="match: GSS: Em:AQ815029"  
 61..595  
 /note="match: GSS: Em:AQ531010"  
 647..944  
 /note="L11PA12 repeat: matches -1422..-1118 of consensus"  
 1339..1734  
 /note="L2 repeat: matches 1652..2042 of consensus"  
 2037..2348  
 /note="AluX repeat: matches 1..312 of consensus"  
 2389..2697  
 /note="AluX repeat: matches 2..310 of consensus"  
 2939..3074  
 /note="FLAM\_C repeat: matches 12..141 of consensus"  
 3197..3411  
 /note="MER58A repeat: matches 1..224 of consensus"  
 3518..3821  
 /note="AluX repeat: matches 1..300 of consensus"  
 4024..4580  
 /note="L2 repeat: matches 2210..2750 of consensus"  
 4626..4766  
 /note="L2 repeat: matches 2240..2383 of consensus"  
 4768..5076  
 /note="AluX repeat: matches 1..312 of consensus"  
 5475..5997  
 /note="L1MCB repeat: matches 102..831 of consensus"  
 5994..6157  
 /note="L1M4 repeat: matches 4140..4298 of consensus"  
 6188..6294  
 /note="FLAM\_C repeat: matches 1..133 of consensus"  
 6295..6750  
 /note="L1M4 repeat: matches 4298..4741 of consensus"  
 6751..7063  
 /note="AluY repeat: matches 1..311 of consensus"  
 7064..7628  
 /note="L1M4 repeat: matches 4741..5310 of consensus"  
 7629..7915  
 /note="AluX repeat: matches 1..284 of consensus"  
 7916..8047  
 /note="L1M4 repeat: matches 5310..5437 of consensus"  
 8054..8180



```
/note="FLAM_C repeat: matches 2. .140 of consensus"
repeat_region 8182. .8411
/note="LIPB7 repeat: matches 5432. .5661 of consensus"
repeat_region 8412. .8717
/note="AluSg repeat: matches 1. .306 of consensus"
repeat_region 8718. .8862
/note="LIPB7 repeat: matches 5661. .5805 of consensus"
repeat_region 8863. .9012
/note="LIPB7 repeat: matches 5992. .6141 of consensus"
repeat_region 9140. .9288
/note="AluJo repeat: matches 1. .151 of consensus"
repeat_region 9292. .9516
/note="LIMB6 repeat: matches 5971. .6182 of consensus"
repeat_region 10088. .10389
/note="AluSg repeat: matches 1. .302 of consensus"
repeat_region 10390. .10697
/note="AluY repeat: matches 1. .308 of consensus"
repeat_region 11104. .11317
/note="MER46A repeat: matches 22. .235 of consensus"
repeat_region 11749. .12047
/note="AluSg repeat: matches 1. .299 of consensus"
repeat_region 12418. .12472
/note="LIMB2 repeat: matches 3968. .4023 of consensus"
repeat_region 12473. .12772
/note="AluSg repeat: matches 1. .299 of consensus"
repeat_region 12773. .13104
/note="LIMB2 repeat: matches 4023. .4362 of consensus"
repeat_region 13105. .13408
/note="AluSg repeat: matches 1. .303 of consensus"
repeat_region 13409. .13722
/note="AluY repeat: matches 1. .310 of consensus"
repeat_region 13410. .14039
/note="LIMB2 repeat: matches 4362. .4638 of consensus"
repeat_region 14040. .14344
/note="AluSx repeat: matches 1. .300 of consensus"
repeat_region 14345. .14683
/note="LIMB2 repeat: matches 4638. .4985 of consensus"
repeat_region 14745. .15005
/note="AluSp repeat: matches 1. .291 of consensus"
/note="match: GSS: Em:AQ536820"
/note="match: GSS: Em:AQ536820"
/note="match: GSS: Em:AQ192026"
/note="match: GSS: Em:AQ192026"
/note="AluJo repeat: matches 1. .134 of consensus"
repeat_region 15310. .15539
/note="LIMB2 repeat: matches 5042. .5257 of consensus"
repeat_region 15540. .15828
/note="AluJo repeat: matches 3. .291 of consensus"
repeat_region 15829. .16547
/note="LIMB2 repeat: matches 5257. .6013 of consensus"
repeat_region 16558. .16666
/note="LITB repeat: matches 2. .121 of consensus"
repeat_region 16675. .16979
/note="AluSx repeat: matches 1. .306 of consensus"
repeat_region 16994. .17094
/note="LIMB2 repeat: matches 6069. .6168 of consensus"
/note="match: GSS: Em:AQ076012"
/note="match: GSS: Em:AQ076012"
/note="Alu repeat: matches 2. .62 of consensus"
repeat_region 18034. .18072
/note="MER4D repeat: matches 564. .601 of consensus"
repeat_region 18073. .18375
/note="AluJo repeat: matches 2. .301 of consensus"
repeat_region 18376. .18422
/note="MER4D repeat: matches 601. .651 of consensus"
repeat_region 18490. .18634
/note="LITB29 repeat: matches 53. .198 of consensus"
repeat_region 18635. .18895
/note="MER41B repeat: matches 1. .271 of consensus"
repeat_region 18947. .19130
/note="AluSg/x repeat: matches 109. .292 of consensus"
```

```
repeat_region 19128. .19305
/note="AluSg repeat: matches 140. .313 of consensus"
repeat_region 19306. .19377
/note="36 copies 2 mer aa 75% conserved"
repeat_region 19723. .19808
/note="LITB repeat: matches 1. .73 of consensus"
repeat_region 19826. .19975
/note="AluJo repeat: matches 137. .286 of consensus"
repeat_region 19976. .20597
/note="LITB repeat: matches 60. .689 of consensus"
/note="match: GSS: Em:AQ891390"
/note="match: GSS: Em:AQ891390"
/note="AluJo repeat: matches 1. .127 of consensus"
repeat_region 21046. .21335
/note="AluSx repeat: matches 6. .297 of consensus"
repeat_region 21336. .21501
/note="AluJo repeat: matches 127. .299 of consensus"
repeat_region 22184. .22478
/note="AluSx repeat: matches 1. .295 of consensus"
repeat_region 22687. .22768
/note="AluSg repeat: matches 2. .84 of consensus"
repeat_region 23426. .23535
/note="HERVL repeat: matches 2. .98 of consensus"
repeat_region 23536. .23826
/note="AluSg repeat: matches 1. .294 of consensus"
repeat_region 23827. .23869
/note="HERVL repeat: matches 98. .143 of consensus"
repeat_region 23932. .24283
/note="HERVL repeat: matches 4267. .4622 of consensus"
repeat_region 24284. .24540
/note="AluSx repeat: matches 1. .294 of consensus"
repeat_region 24541. .24787

Query Match 44.9%; Score 519.2; DB 9; Length 172079;
Best Local Similarity 84.0%; Pred. No. 1.7e-83;
Matches 675; Conservative 0; Mismatches 58; Indels 71; Gaps 5;

QY 91 GGAGGCTGAGACGACGCTGCCATTAATGCGAGTGAAGAAATTAATGAGACTATCCA 150
DB 36364 GGAGTGAAGGCGGTGATGACGATGAGCGAGTGAAGAAATTAATGAGACTATCCA 36305

QY 151 GTGAATTCACAGAGATTTTCAGAGCTTGAAGAAATTCATTTGCTGCTGATGATG 210
DB 36304 GTGAATTCACAGATTTTCAGAGCTTGAAGAAATTCATTTGCTGCTGATGATG 36245

QY 211 CTGAAGACCATGATGCTGTTCTTGAAGAAATGATGTTGCGAGAAATTCACCTTGA 270
DB 36244 CTGAAGACCATGATGCTGTTCTTGAAGAAATGATGTTGCGAGAAATTCACCTTGA 36185

QY 271 CAAGCAAAAGTGATTTGCTTTCGATACACATTAATTCATGTTTGGCTTATTTG 330
DB 36184 CAAGCAAAAGTGATTTGCTTTCGATACACATTAATTCATGTTTGGCTTATTTG 36125

QY 331 GCACCCAGAGATTAATCTTAAGAAACATCCAGTAAACAGAAATTCAGAAATTCACA 330
DB 36124 GCACCCAGAGATTAATCTTAAGAAACATCCAGTAAACAGAAATTCAGAAATTCACA 36065

QY 391 GTATATATGAACAGATCAAGGAAATATACAGACCAAGAAAGGCTGCAAGCTGACAGA 450
DB 36064 GTATATATGAACAGATCAAGGAAATATACAGACCAAGAAAGGCTGCAAGCTGACAGA 36005

QY 451 GGTCAGCTTCAAGATTTGTAAATAAAATGCGCTTGGAAACCAAAATCGAAATTCATCA 510
DB 36004 GGTCAGCTTCAAGATTTGTAAATAAAATGCGCTTGGAAACCAAAATTCATTC 35945

QY 511 AAAGTCCCAATTAAGAAAGAAAGTAAAGTAACTTTTGGTTTGAATGATACATATTC 570
DB 35944 AAAGTCCCAATTAAGAAAGAAAGTAAAGTAACTTTTGGTTTGAATGATACATATTC 35885

QY 571 AAAAAGTACA-----TTAATATGT 589
DB 35884 AAAAAGTACACTCTTTTTCCTCCACAAAGTAATCTGTGACAGACCAAGTTTAATGT 35825
```

QY 590 AATCAGT-----AATGTAAAGTAAATAC-TTCCTC 623  
 DB 35824 ATTCTTGTGAATTAATATGTAATTAAGTAAGTAAGTAAAGTAAATCTTTCCTC 35765  
 QY 624 TCCTAAGATCATTAATCTTTATTTAGTATGACAC--TGAGATTTTAACTGTATATATTA 681  
 DB 35764 TCCAAAGTGTATCTTTATTTATGATTTGACATATGAGATTTTAACTGTATATATTT 35705  
 QY 682 TATATTAATTAATTAACATCTC--TGATAGACTCTTATTTCTTAAATAGTCACTCT 739  
 DB 35704 TATATTAATTAATTAACATCTCTTTGAAAGAGACTCTTATTTCTTATAGTCACTCT 35645  
 QY 740 TCGAAGTCAATTTTAAAGACCTGTGAATTAATTAAGTAAGTCTTTGTAATATTT 799  
 DB 35644 TCGAGTACATTTCTTTGGAAGTGTGAATTAATTAAGTAAATTTCTTTGAAATATTT 35585  
 QY 800 GTTCTTTTAAATGAATATGAC 823  
 DB 35584 GTACTATTTTCAATGAATATGAC 35561

RESULT 15  
 BC005436 2000 bp mRNA linear ROD 29-JUN-2004  
 LOCUS Mus musculus nuclear DNA binding protein, mRNA (cDNA clone MGC:5983  
 DEFINITION IMAGE:3591682), complete cds.  
 ACCESSION BC005436  
 VERSION BC005436.1 GI:13529391  
 KEYWORDS MGC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, P.S., Wagner, J., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Stachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carrinci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, J., Garcia, A.M., Gay, L.J., Hu, Y., S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Schenck, A., Schein, J.E., Jones, S.J., Skalska, U., Small, D.E.,  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16689-16903 (2002)

TITLE  
 JOURNAL  
 PUBMED  
 REFERENCE  
 AUTHORS  
 JOURNAL

2 (bases 1 to 2000)  
 Strausberg, R.  
 Direct Submission  
 Submitted (27-MAR-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>

FEATURES  
 source  
 1..2000  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="MGC:5983 IMAGE:3591682"  
 /tissue\_type="Mammary tumor metastasized to lung. Tumor  
 arose spontaneously from a senescent normal mammary  
 (clonal) outgrowth infected with the virus MMV."  
 /clone\_1lb="NCI CGAP Lu29"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 1..2000  
 /gene="Cld"  
 /note="synonym: SUN-COR"  
 /db\_xref="locusid:57316"  
 /db\_xref="MGI:1927354"  
 47..472  
 /gene="Cld"  
 /codon\_start=1  
 /product="nuclear DNA-binding protein"  
 /protein\_id="AA05436.1"  
 /db\_xref="GI:13529392"  
 /db\_xref="locusid:57316"  
 /db\_xref="MGI:1927354"  
 /translation="MAGEEEDYDVEIEHESITALESSIGAVDMLKTMVSRBEL  
 OKLDPLKQKLVLSAYTLTSMFVYLTGCVNPREHPKQELERIVYMRVKEITD  
 KKKAKLIDGSAASRPVKAALWPKAKSTPKVANKKSGH"

## gene

## CDS

## ORIGIN

Query Match 40.1%; Score 464; DB 10; Length 2000;  
 Best Local Similarity 71.5%; Pred. No. 2.9e-73;  
 Matches 719; Conservative 0; Mismatches 255; Indels 32; Gaps 7;

QY 106 GAGCTGGCCATTAATGCGAGGTGAAGAAATTAATGAAGCTATCCAGTGAATTCAGAG 165  
 DB 35 GAGTGGTCAATGCGAGGTGAAGAAATTAATGAAGCTATTCCTGGAATTCAGAG 94  
 QY 166 TATTTGTCAAGCTTTGAGAAATTCATTTGCTGTGATGAGATGCTGAAGCATGATG 225  
 DB 95 TCTTTAAGAGCCCTGAGAGCTCCCTGGTGTCTGTGATGATGATGCTGAAGCATGATG 154  
 QY 226 TCTGTTTCTAAGAAAGAGTTGTCGAAGTTGATTCATTCGAACAAAGAAAGTGAT 285  
 DB 155 GCTGTTTCTAAGAAAGAGTTGTCGAAGTTGATTCGAACAAAGAAAGTGAT 214  
 QY 286 TTGTTTCTGATACATTAATTCATTTGTTGGTTTATTTTGGCAACCAAGAGTT 345  
 DB 215 TTAGTTTCTGATACATTCATTAATTCATTTGTTGGTTTATTTTGGCAACCAAGAGTT 274  
 QY 346 AATCTTAAGAAATCATCGTAAACAGAAATGGAAGAAATCAGAGTATATATGAACGA 405  
 DB 275 AATCCCAAGAGATCCAGTGAAGAACTGGAAGAAATCAGAGTCTACATTAACGA 334  
 QY 406 GTCAAGAAATTAACAGACAAAGAAAGCTGCGACGTCGACAGAGTGCAGTTCAAGA 465  
 DB 335 GTTAAAGAAATTAACAGACAAAGAAAGCTGCGACGTCGACAGAGTGCAGTTCAAGA 394  
 QY 466 TTTGTAAAGAAATGCGCTCTGGAACCAAAATCGAAAATGATCAAAAAGTTGCCAATAA 525  
 DB 395 TTTGTCAAGAAATGCGCTCTGGAACCAAAATGATCAAAAAGTTGCCAATAA 454  
 QY 526 GGAAGAAATTAAGTAACTTTTGTGATGATCAATATTCAAAAGTACATTAAT 585

```
Db      455 GGGAAAAGCAACATCTATCTTTGGTTTGATGTACATGTTTCAAAAAGTACA-TCCF 513
QY      586 ATGTAATCACAGTAATATGTAAGCTAAATACCTCCCTCCAAAGATCATTAATCTTATT 645
Db      514 TTTTAATCATTTTACAAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 573
QY      646 GAT-----TAGCATGAGATTTTAACATTTGATATATATATATATATATATATATATAT 700
Db      574 AATTCATGTAATTAATTTACATTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTA 633
QY      701 CTCTGATGAGACTTATTCTTTATATATAGTCACTTGCAAGTACCATTTTAAAGC 760
Db      634 GATTAGTTCCTTGTGTTGATTTTCAATATTAAGCATCATGATGTTTAAATATGTAAGA 693
QY      761 AGCTGGAAG-----TTAAGGAAGTCTCTTTGTAACATTTGTAATTTTAAAT 813
Db      694 TATTCTATTAAGCAGTGTGTAATCCAAATGTTCTGTGTAACATTTGTAATTTGTAAT 753
QY      814 GAATTAATGACCTTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 873
Db      754 GACATGTAATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 813
QY      874 ACTATATGATATTAAGAAAG--CGTGAATGACTTAATGTTCAATTTTCTGTATAGA 930
Db      814 AGTAATGATGAGAAATATACAGTGAATTAATACCACTCTGTTTCTGTGAGTTAGTTCA 873
QY      931 TACTTATCATGTTTTCATATATTTTAGAATTTACTGTTTGTATATTTCAAGTGTGAA 990
Db      874 ACATGTTTCTGTGATTTTCTTTTGTGAGTAATGCTGTGATTAATTTCAAGTGTGAA 933
QY      991 ACTAAAGTTTATGTTGTTGTAATTTTGGCATGTGCTGTATGCTCCATTTTAA 1050
Db      934 ATTGAACCTTAAGGCTGTACTTAATTTT-----CATGTTCCATTTTAA 980
QY      1051 TAAATATCAATTCATTAATTTAGATGGAATTAAGTGTATGT 1096
Db      981 TAAAT--GTTCTCATTAATCT-GATGGAACACGTTGTATGT 1023
```

Search completed: January 4, 2005, 11:20:31  
Job time : 5289.49 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 4, 2005, 08:13:30 : Search time 66 Seconds  
(without alignments)  
766.376 Million cell updates/sec

Title: US-09-701-618A-2

Perfect score: 715  
Sequence: 1 MAGEBINEDPVEIHEYLISA.....LMPEKSKNASKVANKSKSKS 141

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*

1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	715	100.0	141	AAV51024	AAV51024 Human CID
2	715	100.0	141	ADBS9764	ADBS9764 Human Pro
3	671	93.8	141	AAO15405	AAO15405 Mammalian
4	667	93.3	140	AD128083	AD128083 Human nuc
5	654	91.5	141	ADBS9762	ADBS9762 Rat Prote
6	646	90.3	141	AAV51025	AAV51025 Murine Cl
7	644	90.1	140	AAO15406	AAO15406 Mammalian
8	519	72.6	102	AAO1231	AAO1231 Human sec
9	266	37.2	54	ABO54730	ABO54730 Human gen
10	150.5	21.0	217	AAAG49051	AAAG49051 Arabidops
11	150.5	21.0	217	AAAG24127	AAAG24127 Arabidops
12	150.5	21.0	253	AAAG49050	AAAG49050 Arabidops
13	150.5	21.0	256	AAAG24126	AAAG24126 Arabidops
14	141.5	19.8	177	AAAG24128	AAAG24128 Arabidops
15	141.5	19.8	177	AAAG49052	AAAG49052 Arabidops
16	126.5	17.7	159	ABBB64082	ABBB64082 Drosophi
17	119.5	16.7	232	ADBP9133	ADBP9133 C. albica
18	110	15.4	184	ABBS5602	ABBS5602 Protein s
19	110	15.4	184	ADK64282	ADK64282 Disease t
20	85	11.9	151	ADH88594	ADH88594 Enterococ
21	85	11.9	1294	ABBS6502	ABBS6502 Enterococ
22	83	11.6	1249	ABU44466	ABU44466 Protein e
23	82	11.4	345	AAAG23880	AAAG23880 Arabidops
24	81.5	11.4	295	ADH85852	ADH85852 Enterococ
25	81.5	11.4	307	ADC97413	ADC97413 E. faeciu

26	81.5	11.4	904	AAV16123	AAV16123 Mycoplasma
27	80.5	11.3	675	ABU43960	ABU43960 Protein e
28	75	11.0	479	AAV92242	AAV92242 Human can
29	79	11.0	479	ABO05690	ABO05690 Human nuc
30	79	11.0	479	ABBB83498	ABBB83498 Large hum
31	79	11.0	479	ABBP64721	ABBP64721 Human pro
32	79	11.0	479	ADJ69180	ADJ69180 Human hea
33	79	11.0	948	ABBS61027	ABBS61027 Drosophi
34	78.5	11.0	152	ADN46815	ADN46815 Thermococ
35	78.5	11.0	508	ABU36173	ABU36173 Protein e
36	78	10.9	703	AAV92662	AAV92662 Mutant mu
37	78	10.9	761	AAV92660	AAV92660 Mutant mu
38	78	10.9	923	ABBS9510	ABBS9510 Drosophi
39	77.5	10.8	1094	ABP73717	ABP73717 Candida a
40	77.5	10.8	2285	AAW98149	AAW98149 Bacillus
41	77	10.8	209	ABM67922	ABM67922 Phototrab
42	77	10.8	465	AAU34291	AAU34291 Staphyloc
43	77	10.8	466	AAW22467	AAW22467 Staphyloc
44	77	10.8	466	AAW58529	AAW58529 Staphyloc
45	77	10.8	466	AAV91053	AAV91053 Staphyloc

## ALIGNMENTS

RESULT 1	
AAV51024	AAV51024 standard; protein; 141 AA.
AC	AAV51024;
XX	
DT	17-MAR-2000 (first entry)
XX	
DE	Human CID protein.
XX	
KM	CID; human; apoptosis; tumour; gene therapy; treatment.
XX	
OS	Homo sapiens.
XX	
PN	DE19824811-A1.
XX	
PD	09-DEC-1999.
XX	
PF	03-JUN-1998; 98DE-01024811.
XX	
PR	03-JUN-1998; 98DE-01024811.
XX	
PA	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI	Rothbarth K, Stammer H, Werner D;
XX	
DR	WPI; 2000-063506/06.
XX	
PT	N-PSDB; AA243927.
XX	
PS	Inducing apoptosis by overexpressing the CID gene, particularly for treating tumours.
XX	
PS	Claim 3; Fig 1; 10pp; German.
XX	
CC	This invention describes a novel method for inducing apoptosis which comprises overexpressing the CID gene (I). The method is particularly used to treat tumors and can also be used in gene therapy. The method has no side effects on normal cells (contrast known methods of inducing apoptosis such as cytotoxins and radiation), and may be effective on cells resistant to conventional treatments. Overexpression of (I) is sufficient itself to induce apoptosis but the effect may be increased when used in combination with other anti-tumor methods. When cells transfected with (I) undergo apoptosis, they release factors that kill neighboring, non-transfected cells (bystander effect). This sequence represents the human CID protein described in the method of the invention
XX	
SQ	Sequence 141 AA;

Query Match 100.0%; Score 715; DB 3; Length 141;  
Best Local Similarity 100.0%; Pred. No. 1.9e-66;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGEINEDYVEIHEYLAFENSIGAVDEMLKTMMSVSRENELQKLDPLEQAKVDLVA 60  
DB 1 MAGEINEDYVEIHEYLAFENSIGAVDEMLKTMMSVSRENELQKLDPLEQAKVDLVA 60  
QY 61 YTLNSMFVYLATQGVNPKHPVKQELERIRVYNNRVKXITDKKXGKLDGGAASRFVN 120  
DB 61 YTLNSMFVYLATQGVNPKHPVKQELERIRVYNNRVKXITDKKXGKLDGGAASRFVN 120  
QY 121 ALWEPKSKNASKVANKGSKS 141  
DB 121 ALWEPKSKNASKVANKGSKS 141

RESULT 2  
ADES9764 standard; protein; 141 AA.  
ADES9764;  
29-JAN-2004 (first entry)  
Human Protein NP\_006324, SEQ ID NO 5660.  
Human; pain; neuronal tissue; gene therapy;  
spinal segmental nerve injury; chronic constriction injury; CCI;  
spared nerve injury; SNI; Chung.  
Homo sapiens.  
MO2003016475-A2.  
27-FEB-2003.  
14-AUG-2002; 2002WO-US025765.  
14-AUG-2001; 2001US-0312147P.  
01-NOV-2001; 2001US-0346382P.  
26-NOV-2001; 2001US-0333347P.  
(GHEO) GEN HOSPITAL CORP.  
(FARB) BAYER AG.  
Woolf C, D'urso D, Befort K, Costigan M;  
WPI; 2003-268312/26.  
GENBANK; NP\_006324.

New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.  
Claim 1; Page: 1017pp; English.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating

CC pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published\_pcl\_sequences.

Sequence 141 AA:

Query Match 100.0%; Score 715; DB 7; Length 141;  
Best Local Similarity 100.0%; Pred. No. 1.9e-66;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGEINEDYVEIHEYLAFENSIGAVDEMLKTMMSVSRENELQKLDPLEQAKVDLVA 60  
DB 1 MAGEINEDYVEIHEYLAFENSIGAVDEMLKTMMSVSRENELQKLDPLEQAKVDLVA 60  
QY 61 YTLNSMFVYLATQGVNPKHPVKQELERIRVYNNRVKXITDKKXGKLDGGAASRFVN 120  
DB 61 YTLNSMFVYLATQGVNPKHPVKQELERIRVYNNRVKXITDKKXGKLDGGAASRFVN 120  
QY 121 ALWEPKSKNASKVANKGSKS 141  
DB 121 ALWEPKSKNASKVANKGSKS 141

RESULT 3  
AA015405 standard; protein; 141 AA.  
AA015405;  
27-SEP-2002 (first entry)  
Mammalian nuclear receptor cofactor CF7 protein.  
Mammalian; nuclear receptor cofactor; CF7; CF8; metabolism regulation;  
cell homeostasis; cell proliferation; differentiation;  
pathological cellular aberration; cellular defence mechanism.  
Mammalia.  
WO200242322-A2.  
30-MAY-2002.  
21-NOV-2001; 2001WO-EP013548.  
21-NOV-2000; 2000EP-00125524.  
(LION-) LION BIOSCIENCE AG.  
Jackson D, Casari G, Suckow J;  
WPI; 2002-566559/60.  
N-PSDB; AAL43973.

Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators useful for inhibiting cellular function of cofactor and for treating metabolic disorders, immunological indications and hormonal dysfunctions.  
Claim 12; Fig 3; 68pp; English.

The invention comprises the amino acid and coding sequences of two mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and CF8 protein sequences of the invention are useful for screening agents that are capable of inhibiting the cellular function of cofactor CF7 and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes involved in cellular functions, such as: regulation of metabolism and

CC cell homeostasis, cell proliferation and differentiation, pathological  
CC cellular aberrations, or cellular defence mechanisms. The present amino  
CC acid sequence represents the mammalian nuclear receptor cofactor CF7  
CC protein  
CC  
CC  
CC Sequence 141 AA,  
SQ

Query Match	93.8%	DB 5:	length	141
Best Local	93.6%	Pred. No. 7.6e-62		
Matches 132	3	Mismatches	6	Indels 0
				Gaps 0

[illegible]

RESULT 4  
ADI28083  
ID ADI28083 standard; protein; 140 AA

AC	ADI28083;
XX	
DT	22-APR-2004 (first entry)
XX	
DE	Human nuclear receptor cofactor C/EBP protein

KW nuclear receptor cofactor; CF6; cellular function inhibition;  
KW metabolic disorder; immunological indication; hormonal dysfunction,  
KW neurosystemic disease.

OS	Homo sapiens.
XX	
PN	WO200224728-A2.

PD 28-MAR-2002

PF 17-SEP-2001; 2001WO-EP010744

PR 22-SEP-2000; 2000EP-00120722.

PA (LION-) LION BIOSCIENCE AG.

PI	Cabard, G, Jackson L
XX	
DR	WPI; 2002-383179/41.
DR	N-PSDB; ADI28090.

PT Polynucleotide and polypeptide of novel nuclear receptor cofactor useful  
PT for screening drugs regulating cofactor-associated physiological  
PT responses e.g. hormonal dysfunctions.

PS Claim 12; SEQ ID NO 3; 97pp; English.

The invention relates to an isolated polynucleotide encoding a nuclear receptor cofactor also known as C/EBP. The polynucleotide or encoded protein is useful for construction of multiple nuclear receptor cofactor constructs, specifically sequence alignments, preferably for the construction of protein sequence alignments. The protein is useful for screening agents capable of inhibiting the cellular function of the cofactor C/EBP. The polynucleotide is useful for making vectors and for transforming cells, both of which are ultimately useful for production of the C/EBP protein. They are also useful as scientific research tools for developing nucleic acid probes for determining expression levels of the cofactor gene, e.g. to identify diseased or otherwise abnormal states. They are particularly useful for diagnostic purposes to e.g., identify deleted or mutant C/EBP

genes, or their measure expression. They are useful for developing analytical tools such as antisense oligonucleotide for selectively inhibiting expression of the cofactor gene to determine physiological responses. The protein is useful for screening drugs for agonist and antagonist activity, and therefore, for screening for drugs useful in regulating physiological responses associated with the cofactors such as metabolic disorders, immunological indications, hormonal dysfunction, neurodegenerative diseases. The proteins are also useful for developing antibodies for detection of the proteins. The polynucleotide can be used to design primers for a polymerase chain reaction and are also used to model the three-dimensional structure of the protein. This sequence represents the C66 protein. (Note: this sequence appears to be a fragment of the C66 protein as the sequence differs from the C66 protein sequence given in Fig 3 (also designated as SEQ ID NO: 2) of the specification).

Query Match	93.3%	Score 667	DB 5	Length 140
Best Local Similarity	93.6%	Pred. No. 2e-61		
Matches 131; Conservative	3	Mismatches	6	Indels 0
				Gaps 0

[illegible]

RESULT 5  
ADE59762  
ID ADE59762 standard; protein; 141 AA

AC ADE59762;

DT 29-JAN-2004 (first entry)

DE Rat Protein AAH05436, SEQ ID NO 5658.

KM Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury  
KM chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

OS Rattus norvegicus

PN WO2003016475-A2.

PD 27-FEB-2003

PF 14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P  
PR 01-NOV-2001; 2001US-0346382P  
PR 26-NOV-2001; 2001US-0333347P

PA (GEHO ) GEN HOSPITAL CORP  
PA (FARB ) BAYER AG.  
PA

PI Woolf C, D'urbo D, Befort K, Costigan M

DR WPI; 2003-268312/26

XX

PT preparing a medicament for treating pain in an animal.

PS Claim 1; Page; 1017pp; English.

CC The invention discloses a composition comprising two or more isolated rat

DR	N-PSDB; AAZ43928.				
XX					
PT	Inducing apoptosis by overexpressing the CID gene, particularly for				
PT	treating tumors.				
XX					
PS	Claim 3; Fig 2; 10pp; German.				
XX					
CC	This invention describes a novel method for inducing apoptosis which				
CC	comprises overexpressing the CID gene (1). The method is particularly				
CC	used to treat tumors and can also be used in gene therapy. The method has				
CC	no side effects on normal cells (contrast known methods of inducing				
CC	apoptosis such as cytotoxins and radiation), and may be effective on				
CC	cells resistant to conventional treatments. Overexpression of (1) is				
CC	sufficient itself to induce apoptosis but the effect may be increased				
CC	when used in combination with other anti-tumor methods. When cells				
CC	transfected with (1) undergo apoptosis, they release factors that kill				
CC	neighboring, non-transfected cells (bystander effect). This sequence				
CC	represents the murine CID protein described in the method of the				
CC	invention				
XX					
XX					
SO	Sequence 141 AA;				
Query Match	90.3%; Score 646; DB 3; Length 141;				
Best Local Similarity	90.0%; Pred. NO. 3, 1e-59;				
Matches 126; Conservative	7; Mismatches 7; Indels 0; Gaps 0;				
1 MATCHES					

Query Match	91.5%;	Score 654;	DB 7;	Length 141;
Best Local Similarity	90.7%;	Pred. No. 4.5e-60;		
Matches 127; Conservative	8;	Mismatches 5.	Indels 0	

[illegible]

```

RESULT 6
AAAY51025
      AY51025 standard; protein; 141 AA
xx

```

1/-MAR-2000 (first entry)  
Murine C1D protein.

apoptosis; tumour; gene therapy; treatment.

DE19824811-A1.

09-DEC-1999

03-JUN-1998; 98DE-01024811.

98DE-01024811.

DEUTSCHES KREBSFORSCHUNGSZENTRUM.

Werner D; Schlatter H, Schlatter N;

WPI; 2000-063506/06.

2

RESULT 7  
AA015406  
ID AA015406 standard; protein, 140 AA.  
AC AA015406;  
XX  
DT 27-SEP-2002 (first entry)  
DE Mammalian nuclear receptor cofactor C98 protein.  
KW Mammalian; nuclear receptor cofactor; CF7, CF8; metabolism regulation;  
KW cell homeostasis; cell proliferation; differentiation;  
XX pathological cellular aberration; cellular defence mechanism.  
OS Mammalia.  
PN WO200242322-A2.  
XX  
PD 30-MAY-2002.  
XX  
PF 21-NOV-2001; 2001MO-EP013548.  
XX  
PR 21-NOV-2000; 2000EP-00125524.  
XX  
PA (LION-) LION BIOSCIENCE AG.  
XX  
PI Jackson D, Casari G, Suckow J;  
DR WPI: 2002-565559/60.  
DR N-PDB: AAL43975.  
XX

Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulator useful for inhibiting cellular function of cofactor and for treating metabolic disorders, immunological indications and hormonal dysfunction.

immunological indications and hormonal dysfunctions.



XX  
PS Claim 12; Fig 3; 68bp; English.  
XX  
CC The invention comprises the amino acid and coding sequences of two  
CC mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and  
CC CF8 protein sequences of the invention are useful for screening agents  
CC that are capable of inhibiting the cellular function of cofactor CF7  
CC and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes  
CC involved in cellular functions, such as: regulation of metabolism and  
CC cell homeostasis, cell proliferation and differentiation, pathological  
CC cellular aberrations, or cellular defence mechanisms. The present amino  
CC acid sequence represents the mammalian nuclear receptor cofactor CF8  
CC protein  
XX  
SQ Sequence 140 AA;  
XX  
Query Match 90.1%; Score 644; DB 5; Length 140;  
Best Local Similarity 91.4%; Pred. No. 5e-59;  
Matches 128; Conservative 2; Mismatches 10; Indels 0; Gaps 0;  
XX  
QY 2 AGGEINEDYVEIHEYLISAFENSIGAVDEMLKTMMSVSRNELLQKDPLEQAKVDLVSA 61  
DB 1 AAGEINEDYVEIHDYLSAFANSIDAVDEMLKMMSSVRNELLQKDPLEQAKVDLVSA 60  
XX  
QY 62 TLNSMFVYLTATGVPNKEHPVKQELERIRVYNNRVKEITDQKXGKLDGASRFVNA 121  
DB 61 TLNSMFVYLTATGVPNKEHSVKQELERIRVYNNRVKEIADKKAGKLDGASRFVNA 120  
XX  
QY 122 LWEPKSKNASKVANKGSKS 141  
DB 121 LWEPKPKNASKVAVKSKS 140  
XX  
RESULT 8  
AAG01231  
ID AAG01231 standard; protein; 102 AA.  
XX  
AC AAG01231;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein, SEQ ID NO: 5312.  
XX  
KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping.  
XX  
OS Homo sapiens.  
XX  
PN EP1033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-00200610.  
XX  
PR 26-FEB-1999; 99US-0122487P.  
XX  
PA (GEST) GENSET.  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
DR WPI; 2000-500381/45.  
DR N-PSDB; AAC01237.  
XX  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX  
PS Claim 13; SEQ ID NO 5312; 71bp + Sequence Listing; English.  
XX  
CC The present sequence is a polypeptide encoded by one of a large number of  
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different  
CC tissues. EST sequences usually correspond mainly to the 3' untranslated

CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
CC sequences derived from the 5' ends of mRNAs and even in those cases where  
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
CC are also used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. They are used to obtain upstream regulatory sequences  
CC and to design expression and secretion vectors  
XX  
SQ Sequence 102 AA;  
XX  
Query Match 72.6%; Score 519; DB 3; Length 102;  
Best Local Similarity 100.0%; Pred. No. 3.7e-46;  
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 MAGEINEDYVEIHEYLISAFENSIGAVDEMLKTMMSVSRNELLQKDPLEQAKVDLVSA 60  
DB 1 MAGEINEDYVEIHEYLISAFENSIGAVDEMLKTMMSVSRNELLQKDPLEQAKVDLVSA 60  
XX  
QY 61 YTLNSMFVYLTATGVPNKEHPVKQELERIRVYNNRVKEITD 102  
DB 61 YTLNSMFVYLTATGVPNKEHPVKQELERIRVYNNRVKEITD 102  
XX  
RESULT 9  
AB054730  
ID AB054730 standard; protein; 54 AA.  
XX  
AC AB054730;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Human genome derived single exon protein #964.  
XX  
KM Human; gene expression; single exon probe; microarray;  
KW alternative splicing event; genomic alteration.  
XX  
OS Homo sapiens.  
XX  
PN US2003194704-A1.  
XX  
PD 16-OCT-2003.  
XX  
PF 03-APR-2002; 2002US-00029386.  
XX  
PR 03-APR-2002; 2002US-00029386.  
XX  
PA (PENN/) PENN S G.  
PA (RANK/) RANK D R.  
PA (HANZ/) HANZEL D K.  
PI Penn SG, Rank DR, Hanzel DK;  
XX  
DR WPI; 2004-119264/12.  
XX  
XX  
PT New human genome-derived single exon nucleic acid probes useful for human  
PT gene expression analysis, for identifying or characterizing alternative  
PT splicing events, for assessing genomic alterations or as tools for  
PT surveying tissues.  
XX  
PS Claim 45; SEQ ID NO 28364; 80bp; English.  
XX  
XX  
CC The invention relates to a nucleic acid probe for measuring human gene  
CC expression, comprising any of the 27,400 fully defined nucleotide  
CC sequences in the specification, or their complements or fragments, and  
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences  
CC fully defined in the specification. The probe is a single exon probe that  
CC hybridizes under high stringency conditions to a nucleic acid molecule  
CC expressed in human cells or tissues. Also included are a spatially-  
CC addressable set of single exon nucleic acid probes for measuring human  
CC gene expression (comprising a plurality of single exon nucleic acid  
CC probes cited above, where each of the plurality of probes is separately

CC and addressably isolatable or amplifiable from the plurality), a single  
CC exon microarray for measuring human gene expression, a method of  
CC measuring human gene expression, a vector comprising the single exon  
CC probe cited above, an ORF-encoded peptide comprising at least 8  
CC contiguous amino acids of any of the above-mentioned amino acid  
CC sequences (optionally with conservative amino acid substitutions), an  
CC isolated antibody that binds specifically to a peptide cited above,  
CC a customer desiring to measure gene expression, a method of providing  
CC human gene expression data by subcription, and a computer-readable  
CC storage medium which contains a database having a plurality of records  
CC each record including data on the expression of a single exon probe  
CC cited above. The probe, methods and apparatus are useful in gene  
CC expression analysis. The probes may be used as tools for surveying  
CC tissues to detect the presence of expressed messages that contain their  
CC specific exon, or in constructing genome-derived single exon microarrays.  
CC In addition, the probes are used in identifying and characterizing  
CC alternative splicing events, in detecting and characterizing gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe protein of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?docID=20030194704  
XX  
SQ Sequence 54 AA;

Query Match 37.2%; Score 266; DB 8; Length 54;  
Best Local Similarity 96.1%; Pred. No. 4.2e-20;  
Matches 53; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 88 ERIIRVYNNRKYETITDKKAGKIDRGASRVKNALMEPKSKAKYANKGSKS 141  
Db 1 ERIIRVYNNRKYETITDKKAGKIDRGASRVKNALMEPKSKAKYANKGSKS 54

RESULT 10  
AAG49051  
ID AAG49051 standard; protein; 217 AA.  
XX

AC AAG49051;  
XX  
DT 18-OCT-2000 (first entry)  
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 62013.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

OS Arabidopsis thaliana.  
XX

PN EPI033405-A2.  
XX

PD 06-SEP-2000.  
XX

PF 25-FEB-2000; 2000EP-00301439.  
XX

PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-01233160P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 16-APR-1999; 99US-0128714P.  
PR 08-APR-1999; 99US-0129845P.  
PR 16-APR-1999; 99US-0130077P.  
PR 19-APR-1999; 99US-0130449P.  
PR 21-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134376P.  
PR 19-MAY-1999; 99US-0134768P.  
PR 20-MAY-1999; 99US-0134941P.  
PR 21-MAY-1999; 99US-0135124P.  
PR 24-MAY-1999; 99US-0135353P.  
PR 25-MAY-1999; 99US-0135629P.  
PR 27-MAY-1999; 99US-0136021P.  
PR 28-MAY-1999; 99US-0136382P.  
PR 01-JUN-1999; 99US-0136782P.  
PR 03-JUN-1999; 99US-0137222P.  
PR 04-JUN-1999; 99US-0137528P.  
PR 07-JUN-1999; 99US-0137502P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139753P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139817P.  
PR 23-JUN-1999; 99US-0139859P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140921P.  
PR 30-JUN-1999; 99US-0141877P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142055P.  
PR 08-JUL-1999; 99US-0142390P.  
PR 09-JUL-1999; 99US-0142803P.  
PR 12-JUL-1999; 99US-0142920P.  
PR 13-JUL-1999; 99US-0142977P.  
PR 14-JUL-1999; 99US-0143542P.  
PR 15-JUL-1999; 99US-0143624P.  
PR 16-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.

```

PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-014512P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148568P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-015138P.
PR 01-SEP-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.

PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 21.0%; Score 150.5; DB 3; Length 217;
Best Local Similarity 27.6%; Pred. No. 2.8e-07;
Matches 37; Conservative 35; Mismatches 47; Indels 15; Gaps 2;

QY 22 ENSIGAVDE-----MLKTMMSVSRNELLQKDPLEQAKVDVSATLNSMFWVL 71
DB 18 ESAIEAVNQTFLAYLKEIKPOLQEMTLAEPEVLAAMQPLQAKTMHLAEATTTLYELRL 77
QY 72 ATGVNPKHEPVKQSELRIRVYNNRVKYEITDKKK-----AGKLDRCASRFVKALWEPR 126
DB 78 RCTGVDPDDHRVSEIERIVNVEKFKVCYDCKGKPLRPPTTVNLRQAATFIHSLPDLT 137
QY 127 SKNASKVANKGSK 140
DB 138 STOKSIRDLKSGE 151

RESULT 11
AAG24127
ID AAG24127 standard; protein; 217 AA.
XX
AC AAG24127;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 27683.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.

```

PR 23-APR-1999; 99US-0130691P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0133256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135622P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 21-JUN-1999; 99US-0139763P.  
PR 22-JUN-1999; 99US-0139817P.  
PR 23-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 24-JUN-1999; 99US-0140354P.  
PR 28-JUN-1999; 99US-0140695P.  
PR 29-JUN-1999; 99US-0140823P.  
PR 30-JUN-1999; 99US-0140991P.  
PR 01-JUL-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141847P.  
PR 02-JUL-1999; 99US-0142154P.  
PR 06-JUL-1999; 99US-0142055P.  
PR 08-JUL-1999; 99US-0142390P.  
PR 09-JUL-1999; 99US-0142803P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144311P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 23-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 26-JUL-1999; 99US-0145218P.  
PR 26-JUL-1999; 99US-0145224P.  
PR 27-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 02-AUG-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0146389P.  
PR 04-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 09-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 10-AUG-1999; 99US-0147935P.  
PR 11-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-014819P.  
PR 13-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 16-AUG-1999; 99US-0148684P.  
PR 17-AUG-1999; 99US-0149368P.  
PR 18-AUG-1999; 99US-0149175P.  
PR 20-AUG-1999; 99US-0148425P.  
PR 20-AUG-1999; 99US-0148722P.  
PR 20-AUG-1999; 99US-0147723P.  
PR 23-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 25-AUG-1999; 99US-0149930P.  
PR 26-AUG-1999; 99US-0150566P.  
PR 27-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 31-AUG-1999; 99US-0151303P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154018P.  
PR 20-SEP-1999; 99US-0154039P.  
PR 22-SEP-1999; 99US-0154779P.  
PR 23-SEP-1999; 99US-0155139P.  
PR 24-SEP-1999; 99US-0155486P.  
PR 28-SEP-1999; 99US-0155659P.  
PR 29-SEP-1999; 99US-0156458P.  
PR 04-OCT-1999; 99US-0156596P.  
PR 05-OCT-1999; 99US-0157117P.  
PR 06-OCT-1999; 99US-0157753P.  
PR 07-OCT-1999; 99US-0157865P.  
PR 08-OCT-1999; 99US-0158029P.  
PR 12-OCT-1999; 99US-0158329P.  
PR 13-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0158293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 14-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.

PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 21.0%; Score 150.5; DB 3; Length 217;  
Best Local Similarity 27.6%; Pred. No. 2.8e-07;  
Matches 37; Conservative 35; Mismatches 47; Indels 15; Gaps 2;

QY 22 ENSIGAVDE-----MKTMMSVSRNELLQKLDPLEQAKVDLSAYTLNMEKVVYL 71

DB 18 ESAIEAVNQTLAVLKEIKLPQLEOMLTIAEPVLAAMQPLQRAKTMHMLAEATTLVETLRL 77

QY 72 ATGCVNPKHPVKGQELERIRVYNNRVKEIRDKKK---AGKLDRGASRFVKNALWEPR 126

DB 78 RCTGVDPDDHVRVSEIRIVNREKFCQCVDSKSGPLRPTTVLNROKATRPFEHSLEPDLT 137

QY 127 SKNASKVANKGSK 140

DB 138 STOKSIRDSKGE 151

RESULT 12  
AAG49050

ID AAG49050 standard; protein; 253 AA.

XX AAG49050;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 62012.

KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridization assay; genetic mapping; gene expression control; promoter;  
termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0132487P.

PR 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134321P.

PR 14-MAY-1999; 99US-0134370P.

PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.

PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.

PR 24-MAY-1999; 99US-0135529P.

PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137528P.

PR 04-JUN-1999; 99US-0137502P.

PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.

PR 14-JUN-1999; 99US-0138847P.

PR 16-JUN-1999; 99US-0139119P.

PR 16-JUN-1999; 99US-0139452P.

PR 16-JUN-1999; 99US-0139453P.

PR 17-JUN-1999; 99US-0139492P.

PR 18-JUN-1999; 99US-0139454P.

PR 18-JUN-1999; 99US-0139455P.

PR 18-JUN-1999; 99US-0139456P.

PR 18-JUN-1999; 99US-0139457P.

PR 18-JUN-1999; 99US-0139458P.

PR 18-JUN-1999; 99US-0139459P.

PR 18-JUN-1999; 99US-0139460P.

PR 18-JUN-1999; 99US-0139461P.

PR 18-JUN-1999; 99US-0139462P.

PR 18-JUN-1999; 99US-0139463P.

PR 18-JUN-1999; 99US-0139750P.

PR 18-JUN-1999; 99US-0139763P.

PR 21-JUN-1999; 99US-0139817P.

PR 22-JUN-1999; 99US-0139899P.

PR 23-JUN-1999; 99US-0140353P.

PR 23-JUN-1999; 99US-0140354P.

PR 24-JUN-1999; 99US-0140695P.

PR 28-JUN-1999; 99US-0140823P.

PR 29-JUN-1999; 99US-0140991P.

PR 30-JUN-1999; 99US-0141287P.

PR 01-JUL-1999; 99US-0141842P.

PR 01-JUL-1999; 99US-0142154P.

PR 02-JUL-1999; 99US-0142055P.

PR 06-JUL-1999; 99US-0142390P.

PR 08-JUL-1999; 99US-0142803P.

PR 09-JUL-1999; 99US-0142920P.

PR 12-JUL-1999; 99US-0142977P.

PR 13-JUL-1999; 99US-0143542P.

PR 14-JUL-1999; 99US-0143624P.

PR 15-JUL-1999; 99US-0144005P.

PR 16-JUL-1999; 99US-0144085P.

PR 16-JUL-1999; 99US-0144086P.

PR 19-JUL-1999; 99US-0144325P.

PR 19-JUL-1999; 99US-0144331P.

PR 19-JUL-1999; 99US-0144332P.

PR 19-JUL-1999; 99US-0144333P.

PR 19-JUL-1999; 99US-0144334P.

PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.

[illegible]

PR 21-APR-1999; 99US-0130449P.  
 PR 23-APR-1999; 99US-0130510P.  
 PR 28-APR-1999; 99US-0130891P.  
 PR 28-APR-1999; 99US-0131449P.  
 PR 30-APR-1999; 99US-0132048P.  
 PR 30-APR-1999; 99US-0132407P.  
 PR 04-MAY-1999; 99US-0132484P.  
 PR 05-MAY-1999; 99US-0132485P.  
 PR 06-MAY-1999; 99US-0132486P.  
 PR 06-MAY-1999; 99US-0132487P.  
 PR 07-MAY-1999; 99US-0132863P.  
 PR 11-MAY-1999; 99US-0134256P.  
 PR 14-MAY-1999; 99US-0134218P.  
 PR 14-MAY-1999; 99US-0134219P.  
 PR 14-MAY-1999; 99US-0134221P.  
 PR 14-MAY-1999; 99US-0134370P.  
 PR 18-MAY-1999; 99US-0134768P.  
 PR 19-MAY-1999; 99US-0134941P.  
 PR 20-MAY-1999; 99US-0135124P.  
 PR 21-MAY-1999; 99US-0135353P.  
 PR 24-MAY-1999; 99US-0135629P.  
 PR 25-MAY-1999; 99US-0136021P.  
 PR 27-MAY-1999; 99US-0136392P.  
 PR 28-MAY-1999; 99US-0136782P.  
 PR 01-JUN-1999; 99US-0137222P.  
 PR 03-JUN-1999; 99US-0137528P.  
 PR 04-JUN-1999; 99US-0137502P.  
 PR 07-JUN-1999; 99US-0137724P.  
 PR 08-JUN-1999; 99US-0138094P.  
 PR 10-JUN-1999; 99US-0138540P.  
 PR 10-JUN-1999; 99US-0138847P.  
 PR 14-JUN-1999; 99US-0139119P.  
 PR 16-JUN-1999; 99US-0139452P.  
 PR 16-JUN-1999; 99US-0139453P.  
 PR 17-JUN-1999; 99US-0139492P.  
 PR 18-JUN-1999; 99US-0139454P.  
 PR 18-JUN-1999; 99US-0139455P.  
 PR 18-JUN-1999; 99US-0139456P.  
 PR 18-JUN-1999; 99US-0139457P.  
 PR 18-JUN-1999; 99US-0139458P.  
 PR 18-JUN-1999; 99US-0139459P.  
 PR 18-JUN-1999; 99US-0139460P.  
 PR 18-JUN-1999; 99US-0139461P.  
 PR 18-JUN-1999; 99US-0139462P.  
 PR 18-JUN-1999; 99US-0139463P.  
 PR 18-JUN-1999; 99US-0139750P.  
 PR 18-JUN-1999; 99US-0139763P.  
 PR 21-JUN-1999; 99US-0139817P.  
 PR 22-JUN-1999; 99US-0139899P.  
 PR 23-JUN-1999; 99US-0140353P.  
 PR 23-JUN-1999; 99US-0140354P.  
 PR 24-JUN-1999; 99US-0140659P.  
 PR 28-JUN-1999; 99US-0140823P.  
 PR 28-JUN-1999; 99US-0140991P.  
 PR 30-JUN-1999; 99US-0141287P.  
 PR 01-JUL-1999; 99US-0141842P.  
 PR 01-JUL-1999; 99US-0142154P.  
 PR 02-JUL-1999; 99US-0142055P.  
 PR 06-JUL-1999; 99US-0142390P.  
 PR 08-JUL-1999; 99US-0142803P.  
 PR 09-JUL-1999; 99US-0142820P.  
 PR 12-JUL-1999; 99US-0142877P.  
 PR 13-JUL-1999; 99US-0143342P.  
 PR 14-JUL-1999; 99US-0143624P.  
 PR 15-JUL-1999; 99US-0144005P.  
 PR 16-JUL-1999; 99US-0144085P.  
 PR 16-JUL-1999; 99US-0144086P.  
 PR 19-JUL-1999; 99US-0144325P.  
 PR 19-JUL-1999; 99US-0144331P.  
 PR 19-JUL-1999; 99US-0144332P.  
 PR 19-JUL-1999; 99US-0144333P.  
 PR 19-JUL-1999; 99US-0144334P.  
 PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.  
 PR 20-JUL-1999; 99US-0144632P.  
 PR 20-JUL-1999; 99US-0144884P.  
 PR 21-JUL-1999; 99US-0144814P.  
 PR 21-JUL-1999; 99US-0145086P.  
 PR 21-JUL-1999; 99US-0145087P.  
 PR 22-JUL-1999; 99US-0145088P.  
 PR 22-JUL-1999; 99US-0145089P.  
 PR 22-JUL-1999; 99US-0145089P.  
 PR 22-JUL-1999; 99US-0145192P.  
 PR 23-JUL-1999; 99US-0145145P.  
 PR 23-JUL-1999; 99US-0145218P.  
 PR 23-JUL-1999; 99US-0145224P.  
 PR 26-JUL-1999; 99US-0145276P.  
 PR 27-JUL-1999; 99US-0145913P.  
 PR 27-JUL-1999; 99US-0145918P.  
 PR 28-JUL-1999; 99US-0145919P.  
 PR 28-JUL-1999; 99US-0145951P.  
 PR 02-AUG-1999; 99US-0146386P.  
 PR 02-AUG-1999; 99US-0146388P.  
 PR 02-AUG-1999; 99US-0146389P.  
 PR 03-AUG-1999; 99US-0147038P.  
 PR 04-AUG-1999; 99US-0147204P.  
 PR 04-AUG-1999; 99US-0147302P.  
 PR 05-AUG-1999; 99US-0147192P.  
 PR 05-AUG-1999; 99US-0147260P.  
 PR 06-AUG-1999; 99US-0147703P.  
 PR 06-AUG-1999; 99US-0147416P.  
 PR 09-AUG-1999; 99US-0147493P.  
 PR 09-AUG-1999; 99US-0147935P.  
 PR 10-AUG-1999; 99US-0148171P.  
 PR 11-AUG-1999; 99US-0148319P.  
 PR 12-AUG-1999; 99US-0148341P.  
 PR 13-AUG-1999; 99US-0148565P.  
 PR 16-AUG-1999; 99US-0149368P.  
 PR 17-AUG-1999; 99US-0149175P.  
 PR 18-AUG-1999; 99US-0149426P.  
 PR 20-AUG-1999; 99US-0149722P.  
 PR 20-AUG-1999; 99US-0149723P.  
 PR 20-AUG-1999; 99US-0149929P.  
 PR 23-AUG-1999; 99US-0149902P.  
 PR 23-AUG-1999; 99US-0149930P.  
 PR 25-AUG-1999; 99US-0150566P.  
 PR 26-AUG-1999; 99US-0150884P.  
 PR 27-AUG-1999; 99US-0151065P.  
 PR 27-AUG-1999; 99US-0151066P.  
 PR 27-AUG-1999; 99US-0151080P.  
 PR 30-AUG-1999; 99US-0151303P.  
 PR 31-AUG-1999; 99US-015138P.  
 PR 01-SEP-1999; 99US-0151930P.  
 PR 07-SEP-1999; 99US-0152363P.  
 PR 10-SEP-1999; 99US-0153070P.  
 PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154039P.  
 PR 20-SEP-1999; 99US-0154479P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 23-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155659P.  
 PR 28-SEP-1999; 99US-0156458P.  
 PR 29-SEP-1999; 99US-015656P.  
 PR 04-OCT-1999; 99US-0157171P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157865P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158332P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.

```
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160800P.  
PR 22-OCT-1999; 99US-0160811P.  
PR 22-OCT-1999; 99US-0160819P.  
PR 25-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 26-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161924P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.
```

```
Query Match 21.0%; Score 150.5; DB 3; Length 256;  
Best Local Similarity 27.6%; Pred. No. 3.5e-07;  
Matches 37; Conservative 35; Mismatches 47; Indels 15; Gaps 2;
```

```
QY 22 ENSIGAVDE-----MIKTMSVSRNELQKLDPLEQAVDLSAVYTINSMEWYL 71  
DB 57 ESALBAVNOQLAYIKELPQLEQMLTLAEPEVLAAMQPIQRAKMTMLAEATTLYEURL 116  
QY 72 ATQGVNKRERPVKQELERIRVYNNRVRKEITDKK-----AGLIDGASRSFYKNALMPEPK 126  
DB 117 RCTVDDDDRHVRVSEIRVIVYKEFKQCVDSKGPRLPTVLNROAATRFTEHSLPLT 176  
QY 127 SKNASKXANKGK 140  
DB 177 STOKOSIRLDSKE 190
```

```
RESULT 14  
ID AAG24128 standard; protein; 177 AA.  
XX
```

```
AC AAG24128;  
XX  
DT 17-OCT-2000 (first entry)  
XX
```

```
DE Arabidopsis thaliana protein fragment SEQ ID NO: 27684.  
XX
```

```
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX
```

```
OS Arabidopsis thaliana.  
XX
```

```
PN EP1033405-A2.  
XX
```

```
PD 06-SEP-2000.  
XX
```

```
PF 25-FEB-2000; 2000EP-00301439.  
XX
```

```
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.
```

```
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130591P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 07-MAY-1999; 99US-0132486P.  
PR 11-MAY-1999; 99US-0132487P.  
PR 14-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0132864P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135124P.  
PR 24-MAY-1999; 99US-0135262P.  
PR 25-MAY-1999; 99US-0135353P.  
PR 27-MAY-1999; 99US-0136021P.  
PR 28-MAY-1999; 99US-0136392P.  
PR 01-JUN-1999; 99US-0137722P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138447P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 14-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140035P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 24-JUN-1999; 99US-0140354P.  
PR 28-JUN-1999; 99US-0140695P.  
PR 29-JUN-1999; 99US-0140823P.  
PR 30-JUN-1999; 99US-0140991P.  
PR 01-JUL-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142154P.  
PR 06-JUL-1999; 99US-0142055P.  
PR 08-JUL-1999; 99US-0142390P.  
PR 09-JUL-1999; 99US-0142803P.  
PR 12-JUL-1999; 99US-0142920P.  
PR 13-JUL-1999; 99US-0142977P.  
PR 14-JUL-1999; 99US-0143542P.  
PR 15-JUL-1999; 99US-0143542P.  
PR 16-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.
```



	PR	19-JUL-1999;	99US-0144332P.
	PR	20-JUL-1999;	99US-0144352P.
	PR	20-JUL-1999;	99US-0144632P.
	PR	20-JUL-1999;	99US-0144884P.
	PR	21-JUL-1999;	99US-0145086P.
	PR	21-JUL-1999;	99US-0145088P.
	PR	22-JUL-1999;	99US-0145085P.
	PR	22-JUL-1999;	99US-0145087P.
	PR	22-JUL-1999;	99US-0145089P.
	PR	23-JUL-1999;	99US-0145192P.
	PR	23-JUL-1999;	99US-0145145P.
	PR	23-JUL-1999;	99US-0145218P.
	PR	26-JUL-1999;	99US-0145224P.
	PR	27-JUL-1999;	99US-0145276P.
	PR	27-JUL-1999;	99US-0145913P.
	PR	27-JUL-1999;	99US-0145918P.
	PR	27-JUL-1999;	99US-0145919P.
	PR	28-JUL-1999;	99US-0145951P.
	PR	02-AUG-1999;	99US-0146386P.
	PR	02-AUG-1999;	99US-0146388P.
	PR	02-AUG-1999;	99US-0146389P.
	PR	03-AUG-1999;	99US-0147038P.
	PR	04-AUG-1999;	99US-0147204P.
	PR	04-AUG-1999;	99US-0147302P.
	PR	05-AUG-1999;	99US-0147192P.
	PR	05-AUG-1999;	99US-0147260P.
	PR	06-AUG-1999;	99US-0147303P.
	PR	06-AUG-1999;	99US-0147416P.
	PR	09-AUG-1999;	99US-0147493P.
	PR	09-AUG-1999;	99US-0147935P.
	PR	10-AUG-1999;	99US-0148171P.
	PR	11-AUG-1999;	99US-0148319P.
	PR	12-AUG-1999;	99US-0148341P.
	PR	13-AUG-1999;	99US-0148565P.
	PR	13-AUG-1999;	99US-0148684P.
	PR	16-AUG-1999;	99US-0149368P.
	PR	17-AUG-1999;	99US-0149175P.
	PR	18-AUG-1999;	99US-0149426P.
	PR	20-AUG-1999;	99US-0149722P.
	PR	20-AUG-1999;	99US-0149723P.
	PR	20-AUG-1999;	99US-0149829P.
	PR	23-AUG-1999;	99US-0149902P.
	PR	23-AUG-1999;	99US-0149930P.
	PR	25-AUG-1999;	99US-0150566P.
	PR	26-AUG-1999;	99US-0150884P.
	PR	27-AUG-1999;	99US-0151065P.
	PR	27-AUG-1999;	99US-0151066P.
	PR	27-AUG-1999;	99US-0151080P.
	PR	30-AUG-1999;	99US-0151303P.
	PR	31-AUG-1999;	99US-0151438P.
	PR	01-SEP-1999;	99US-0151930P.
	PR	07-SEP-1999;	99US-0152363P.
	PR	10-SEP-1999;	99US-0153070P.
	PR	13-SEP-1999;	99US-0153758P.
	PR	15-SEP-1999;	99US-0154018P.
	PR	16-SEP-1999;	99US-0154039P.
	PR	20-SEP-1999;	99US-0154779P.
	PR	22-SEP-1999;	99US-0155139P.
	PR	23-SEP-1999;	99US-0155486P.
	PR	24-SEP-1999;	99US-0155659P.
	PR	28-SEP-1999;	99US-0156488P.
	PR	29-SEP-1999;	99US-0156569P.
	PR	04-OCT-1999;	99US-0157117P.
	PR	05-OCT-1999;	99US-0157753P.
	PR	06-OCT-1999;	99US-0157655P.
	PR	07-OCT-1999;	99US-0158029P.
	PR	08-OCT-1999;	99US-0158232P.
	PR	12-OCT-1999;	99US-0158369P.
	PR	13-OCT-1999;	99US-0159293P.
	PR	13-OCT-1999;	99US-0159294P.
	PR	13-OCT-1999;	99US-0159295P.
	PR	14-OCT-1999;	99US-0159329P.
	PR	14-OCT-1999;	99US-0159329P.
	PR	14-OCT-1999;	99US-0159330P.
	PR	14-OCT-1999;	99US-0159637P.
	PR	14-OCT-1999;	99US-0159638P.
	PR	18-OCT-1999;	99US-0159584P.
	PR	21-OCT-1999;	99US-0160741P.
	PR	21-OCT-1999;	99US-0160767P.
	PR	21-OCT-1999;	99US-0160768P.
	PR	21-OCT-1999;	99US-0160770P.
	PR	21-OCT-1999;	99US-0160814P.
	PR	21-OCT-1999;	99US-0160815P.
	PR	22-OCT-1999;	99US-0160980P.
	PR	22-OCT-1999;	99US-0160981P.
	PR	22-OCT-1999;	99US-0160989P.
	PR	25-OCT-1999;	99US-0161040P.
	PR	25-OCT-1999;	99US-0161405P.
	PR	25-OCT-1999;	99US-0161406P.
	PR	26-OCT-1999;	99US-0161359P.
	PR	26-OCT-1999;	99US-0161360P.</

PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131448P.  
PR 30-APR-1999; 99US-0132004P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 24-JUN-1999; 99US-0140354P.  
PR 28-JUN-1999; 99US-0140695P.  
PR 29-JUN-1999; 99US-0140823P.  
PR 30-JUN-1999; 99US-0140991P.  
PR 01-JUL-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142154P.  
PR 06-JUL-1999; 99US-0142055P.  
PR 08-JUL-1999; 99US-0142390P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 23-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 26-JUL-1999; 99US-0145224P.  
PR 27-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 02-AUG-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0146389P.  
PR 04-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 09-AUG-1999; 99US-0147933P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148119P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 23-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 25-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 31-AUG-1999; 99US-0151303P.  
PR 01-SEP-1999; 99US-0151438P.  
PR 07-SEP-1999; 99US-0151930P.  
PR 10-SEP-1999; 99US-0152363P.  
PR 13-SEP-1999; 99US-0153070P.  
PR 15-SEP-1999; 99US-0153758P.  
PR 16-SEP-1999; 99US-0154018P.  
PR 20-SEP-1999; 99US-0154039P.  
PR 22-SEP-1999; 99US-0154779P.  
PR 23-SEP-1999; 99US-0155139P.  
PR 24-SEP-1999; 99US-0155486P.  
PR 28-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 05-OCT-1999; 99US-0157117P.  
PR 06-OCT-1999; 99US-0157753P.  
PR 07-OCT-1999; 99US-0157865P.  
PR 08-OCT-1999; 99US-0158029P.  
PR 12-OCT-1999; 99US-0158232P.  
PR 13-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 14-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.

PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161922P.  
 PR 28-OCT-1999; 99US-0161993P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 19.8%; Score 141.5; DB 3; Length 177;  
 Best Local Similarity 28.8%; Pred. No. 1.9e-06;  
 Matches 32; Conservative 30; Mismatches 44; Indels 5; Gaps 1;

OY 35 MMSVSRNELLOKDPLEQAKVDVLSAYTLNSMFWVYLATOGVNPKEHPVKQELERIVYM 94  
 DB 1 MLTLASPEVLAAMQPLORAKTMHLAEATTTLYELRLRCTGVDPDDHVRKSEIERINVYR 60  
 OY 95 NRVKEITDKKK-----AGKLDGASRFVKNALWEPKSKNASKVANKGSK 140  
 DB 61 EKFOKCVDSQSGPLRPFTTVLNROHATRFIEHSLPDLTSTOKOSIRDSKGE 111

Search completed: January 4, 2005, 13:42:24  
 Job time : 78 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 4, 2005, 07:11:31 : Search time 4255.51 Seconds

(without alignments)  
9898.769 Million cell updates/sec

Title: US-09-701-618a-1

Perfect score: 1156

Sequence: 1 cttccggagagactggagc.....aggctgaataatgttgaataa 1156

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 1821965908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

EST.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	765.2	66.2	988	4	BM923929
2	650.6	56.3	915	7	CN645648
3	616	53.3	993	7	CO647214
4	606.4	52.5	880	4	BG193528
5	596.6	51.6	1014	4	BM451300
6	592.2	51.2	847	4	BG189347
7	590.2	51.1	853	4	BG217959
8	584.6	50.6	849	4	BG185663
9	579.6	50.1	878	4	BI753996
10	578.2	50.0	781	5	BU854138
11	575.4	49.8	887	1	AL540005
12	568	49.1	807	4	BM679939
13	567	49.0	909	1	AI800532
14	564.8	48.9	1209	4	BG213559
15	562.4	48.7	859	4	BG206383
16	551.6	47.7	782	5	BU852693
17	543.4	47.0	879	4	BG210469
18	534.4	46.2	642	5	BM973398
19	532.8	46.1	575	1	AI553826
20	532.8	46.1	579	1	AI004680
21	532.8	46.1	609	6	CA943561
22	532.8	46.1	640	5	BU675203
23	532.8	46.1	653	5	BU628390
24	532.8	46.1	713	2	AM960444

c	25	531.8	46.0	609	1	AI809071	AI809071
c	26	531.4	46.0	806	4	BI836412	BI836412
c	27	531.2	46.0	571	1	AI379715	AI379715
c	28	531.2	46.0	607	1	AI800846	AI800846
c	29	524.4	45.4	642	6	CA314481	CA314481
c	30	524.2	45.3	716	7	CN342947	CN342947
c	31	522.2	45.2	544	2	AM024957	AM024957
c	32	522	45.2	788	1	AV763046	AV763046
c	33	521.8	45.1	571	1	AI123353	AI123353
c	34	521.4	45.1	535	1	AI435863	AI435863
c	35	521.4	45.1	831	1	BG185139	BG185139
c	36	520.8	45.1	539	2	BE857571	BE857571
c	37	520	45.0	548	2	AM18590	AM18590
c	38	515.4	44.6	586	4	BM675402	BM675402
c	39	515.2	44.6	813	5	BU68056	BU68056
c	40	510.6	44.2	654	1	AV686813	AV686813
c	41	510.6	44.2	738	5	BU561215	BU561215
c	42	508.4	44.0	535	2	BF434793	BF434793
c	43	506.2	43.8	963	6	CD515525	CD515525
c	44	499.2	43.2	838	4	BG210470	BG210470
c	45	494.6	42.8	602	1	AV685739	AV685739

## ALIGNMENTS

RESULT 1  
LOCUS BM923929 988 bp mRNA linear EST 12-MAR-2002  
DEFINITION AGENCOURT\_6709679 NIH\_MGC\_116 Homo sapiens cDNA clone IMAGE:5760072  
5', mRNA Sequence.  
BM923929 BM923929.1 GI:19374308

## KEYWORDS

EST.  
Homo sapiens (human)  
Homo sapiens

## REFERENCE

1 (bases 1 to 988)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

## JOURNAL

Contact: Robert Strausberg, Ph.D.  
Email: cgabs@mail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1AM12806 row: 1 column: 01  
High quality sequence stop: 722.

## FEATURES

### source

1..988  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5760072"  
/lab\_host="DH10B"

/clone\_jib="NIH\_MGC\_116"  
/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb. Insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 66.2%; Score 765.2; DB 4; length 988;  
 Best Local Similarity 90.6%; Pred. No. 3.1e-150;  
 Matches 885; Conservative 0; Mismatches 26; Indels 66; Gaps 4;

```

QY 91 GAGGGGTGAGACAGAGCTGGCCATTAATGCGAGGTGAAGAAATTAATGAAGACTATCCA 150
DB 8 GAGGTGGAAGCCGTGAGTCAGCCATTAATGCGAGGTGAAGAAATTAATGAAGACTATCCA 150
QY 151 GTAGAAATTCAGAGATTTTGTTCAGCCGTTTGAAGAAATTCATTTGGTGTGATGATGATG 210
DB 68 GTAGAAATTCAGAGATTTTGTTCAGCCGTTTGAAGAAATTCATTTGGTGTGATGATGATG 210
QY 211 CTGAAGACCATGATGCTGTTTCTAGAAATGATTTTGGAAAGTTGGAAGTTGCAATCTTGA 270
DB 128 CTGAAGACCATGATGCTGTTTCTAGAAATGATTTTGGAAAGTTGGAAGTTGCAATCTTGA 270
QY 271 CAAGCAAAAGTGATTTGTTTGTTCAGATCATTAAATCAATGTTTGGTTTATTTTG 330
DB 188 CAAGCAAAAGTGATTTGTTTGTTCAGATCATTAAATCAATGTTTGGTTTATTTTG 330
QY 331 GCAACCCAGAGATTTAATCTTAAGAAATCCAGTAAACAGAAATTTGGAAGATCAGA 390
DB 248 GCAACCCAGAGATTTAATCTTAAGAAATCCAGTAAACAGAAATTTGGAAGATCAGA 390
QY 391 GTATATATGAACAGATCAGAAATTAACAGACAGAAAGGCTGGCAAGCTGACAGA 450
DB 308 GTATATATGAACAGATCAGAAATTAACAGACAGAAAGGCTGGCAAGCTGACAGA 450
QY 451 GGTGACCTTCAAGATTTGTAAATAATGCTCTGGGAAACCAAAATCCGAAATGATCATCA 510
DB 368 GGTGACCTTCAAGATTTGTAAATAATGCTCTGGGAAACCAAAATCCGAAATGATCATCA 510
QY 511 AAAGTTGCGAATTAAGAAAGAAATTAAGTTTAACTTTTGGTTTGTATGATGATGATG 570
DB 428 AAAGTTGCGAATTAAGAAAGAAATTAAGTTTAACTTTTGGTTTGTATGATGATGATG 570
QY 571 AAAAAGTAC----- 579
DB 488 AAAAAGTACATCTTCCGCCACCCCGCAAAATATCTGTGCAAGGCAAGTTTAA 547
QY 580 -----ATTATATATGTAACA--GTAATATGTAAGCTAAATAC-TTCCTCTC 625
DB 548 TGTGTTTCTTAAATATGTAATTAATGTAATTAATGTAATTAATGTAATTAATGTAATTA 607
QY 626 CAAAGATCATTAATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 685
DB 608 CAAAGATCATTAATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 685
QY 686 TTTATATTTTCACTCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 745
DB 668 TTTATATTTTCACTCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 745
QY 746 TACCAATTTATAGACGTGTAATAATTAAGTAAATGTTCTTTGTAACATTTGTAATCA 805
DB 728 TACCAATTTATAGACGTGTAATAATTAAGTAAATGTTCTTTGTAACATTTGTAATCA 805
QY 806 TTTTAAATGAATATGATCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 865
DB 788 TTTTAAATGAATATGATCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 865
QY 866 CTGTTTCACTATATGATTTTGAAGAAAGGTG-AAAGATTTAAATGTTGATTTTCTG 924
DB 848 CTGTTTCACTATATGATTTTGAAGAAAGGTG-AAAGATTTAAATGTTGATTTTCTG 924
QY 925 TATAGATCTTATGATGATTTTGAAGATTTGATGATGATGATGATGATGATGATGATG 984
DB 908 TATAGATCTTATGATGATTTTGAAGATTTGATGATGATGATGATGATGATGATGATG 984
QY 985 TGTGAATCTAAAGTTT 1001
DB 968 TGTGAATCTAAAGTTT 984

```

RESULT 2  
 CN645648  
 LOCUS  
 DEFINITION  
 ILLUMIGEN MCQ\_24688 katze\_Mp12 Macaca mulatta cDNA clone  
 IBTUM:10464 5' similar to bases 9 to 906 highly similar to human  
 CID (Hs.380759), mRNA sequence.  
 CN645648  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Macaca mulatta (rhesus monkey)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 Cercopithecinae; Macaca.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: C. Magness  
 Illumigen Biotechnologies Inc.  
 2203 Airport Way S, Suite 450, Seattle, WA 98134, USA  
 Tel: 2063780400  
 Fax: 2063780408  
 Email: cmagness@illumigen.com  
 Sequenced on 2004.03.02. 821 020 bases.  
 PCR Primers  
 FORWARD: CCTCATTAAGGGAACAAA  
 BACKWARD: CACTATAGGCGAATGGTA  
 Insert Length: 915 Std Error: 0.00  
 Plate: CL000128 row: A column: 07  
 Seq primer: CCTCATTAAGGGAACAAA  
 POLYA-No.

FEATURES  
 source  
 1..915  
 /location=Qualifiers  
 /organism="Macaca mulatta"  
 /mol\_type="mRNA"  
 /strain="Indian"  
 /db\_xref="taxon:9544"  
 /clone="IBTUM:10464"  
 /sex="male"  
 /dev\_stage="newborn infant"  
 /lab\_host="E. coli SOLR"  
 /clone\_1lb="katze\_Mp12"  
 /note="Organ: Placenta; Vector: Uni-ZAP XR; Site 1: EcoR  
 I; Site 2: Xho I; Created from Stragene ZAP-cDNA  
 Synthesis kit (catalog #200400) and ZAP-CDNA GigaPack III  
 Gold Cloning Kit (catalog #200450)"

## ORIGIN

Query Match 56.3%; Score 650.6; DB 7; length 915;  
 Best Local Similarity 86.9%; Pred. No. 3.6e-126;  
 Matches 792; Conservative 0; Mismatches 44; Indels 75; Gaps 4;

```

QY 100 GAGCAGACGTGGCCATTAATGCGAGGTGAAGAAATTAATGAAGCTATCCGTAATTT 159
DB 5 GGCCCTGAGAGTCAGCCATTAATGCGAGGTGAAGAAATTAATGAAGCTATCCGTAATTT 159
QY 160 CAGAGATTTTTCAGCCGTTTGAAGATTCATTGCTGTGATGATGATGATGATGATGATG 219
DB 65 CAGAGATTTTTCAGCCGTTTGAAGATTCATTGCTGTGATGATGATGATGATGATGATG 219
QY 220 ATGATGCTGTTTCTAGAAATGATGTTTGAAGAGTTTGAATCCACTTGAACAGCAAAA 279
DB 125 ATGATGCTGTTTCTAGAAATGATGTTTGAAGAGTTTGAATCCACTTGAACAGCAAAA 279
QY 280 GTGATTTGTTTCTGCAATCACTTAATTAATGATTTTGGTTTATTTGGCAACTCAA 339
DB 185 GTGATTTGTTTCTGCAATCACTTAATTAATGATTTTGGTTTATTTGGCAACTCAA 339
QY 340 GAGGTTAATCTTAAGAAACATCCAGTAAACAGAAATTTGGAAGATCGAGTATATATG 399
DB 245 GAGGTTAATCTTAAGAAACATCCAGTAAACAGAAATTTGGAAGATCGAGTATATATG 399

```

```
QY 400 AACAGAGTCAGAGAAATTAACAGACAGAAAAAGGCTGGCAGCTGACAGAGTGCAGCT 459
DB 305 AACAGAGTCAGAGAAATTAACAGACAGAAAAAGGCTGGCAGCTGACAGAGTGCAGCT 364
QY 460 TCAGAGTTTGTAAAAAATGCGCTCTGGGAAACCAAAATCGAAAAATGATCAAAAGTTGCC 519
DB 365 TCAGAGTTTGTAAAAAATGCGCTCTGGGAAACCAAAATGATCAAAAGTTGCC 424
QY 520 AATTAAGAAAAAGTAAAGTAACTTTTGGTTTGTATGATACATATTCAAAAAGTAC 579
DB 425 AATTAAGAAAAAGTAAAGTAACTTTTGGTTTGTATGATACATATTCAAAAAGTAC 484
QY 580 ATTAAATATGTAATAC----- 595
DB 485 ATCTTTTTCCTCCCTCCAGAAATTAATCTGTGCAGAGCAAGTTAAATGTAATT 544
QY 596 -----AGTAATATGTAAGCTAAATAC--TTCTCTCCA 627
DB 545 CTATTGAAATTAATATGTAATTAATTAACAGTAATATGTAAGCTAAATACCTTTCTCTCCA 604
QY 628 AAGATCATTAATCTTATGATGATGACAGAGATTTTAACATTTGATATATTAATATTT 687
DB 605 AAGATGTAATCTTATGATGATGACAGAGATTTTAACATTTGATATATATATATTT 664
QY 688 TATTAATTTACATCTCTT--GATGAGACTCTTATTTCTTATATATAGCTCAGCTTGCAG 745
DB 665 TATTAATTTACATCTCTTGTGATGAGACTCTTATTTCTTATATATAGCTCAGCTTGCAG 724
QY 746 TACCATTTTAAAGCAGCTGTAATTTAAGTAATGTTCTTGTAAACATTTGTACTA 805
DB 725 TACCATTTTAAAGCAGCTGTAATTTAAGTAATGTTCTTGTAAACATTTGTACTA 784
QY 806 TTTTAAATGAATAATGACCTTATGATGATGATCTGATGAGCTGAAATTAATGATACAT 865
DB 785 TTTTAAATGAATAATGACCTTATGATGATGATCTGATGAGCTGAAATTAATGATACAT 844
QY 866 CTGTTTCACTATATGATATTTAAGAAAGCGTG-AAATGACTTAAATGTTCTTTTCTG 924
DB 845 CTGTTTCACTATATGATATTTAAGAAAGCGTG-AAATGACTTAAATGTTCTTTTCTG 904
QY 925 TATGATACCTT 935
DB 905 TAGTATTTT 915
```

```
RESULT 3
CO647214 993 bp mRNA linear EST 23-JUL-2004
LOCUS ILLUMIGEN MQ0 40181 Katze_MMPB2 Macaca mulatta cDNA clone
DEFINITION ILLUMIGEN MQ0 40181 Katze_MMPB2 Macaca mulatta cDNA clone
            CID (Hs.380759), mRNA sequence.
ACCESSION CO647214
VERSION CO647214.1 GI:50568708
SOURCE EST.
KEYWORDS Macaca mulatta (rhesus monkey)
          Macaca mulatta
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
          Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 993)
AUTHORS Katze,M.G., Thomas,M., Korth,M., Iadonato,S.P. and Magness,C.L.
TITLE Large-scale Rhesus Macaque cDNA Sequencing
JOURNAL Unpublished (2003)
COMMENT Contact: C. Magness
          Illumigen Biosciences Inc.
          2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
          Tel: 2063780400
          Fax: 2063780408
          Email: cmagness@illumigen.com
          Sequenced on 2004.05.28. 778 Q20 bases. Library Preparation: Prof.
          Michael Katze Lab at University of Washington DNA Sequencing:
          Illumigen Biosciences Inc. For further information, see
```

```
http://www.macaque.org
PCR Primers
FORWARD: CCTCACTAAGGGAACAAA
BACKWARD: CACTATAGGGCAATTTGGTA
Insert Length: 993 Std Error: 0.00
Plate: C1000339 row: G column: 08
Seq primer: CCTCACTAAGGGAACAAA
POLYA=yes.

FEATURES
    source
        1..993
            /organism="Macaca mulatta"
            /mol_type="mRNA"
            /strain="Indian"
            /db_xref="taxon:9544"
            /clone="IBIUM:23958"
            /sex="male"
            /tissue_type="blood"
            /cell_type="PBMC"
            /dev_stage="adult"
            /lab_host="Electromax DH10B"
            /clone_id="Katze MMPB2"
            /note="Vector: pDONR 222; Site_1: Bam I; Site_2: Bam I;
            Created from Cloneminer cDNA Library Construction kit
            (catalog #18249-029)"

ORIGIN
Query Match 53.3%; Score 616; DB 7; Length 993;
Best Local Similarity 85.5%; Pred. No. 6,7e-119;
Matches 776; Conservative 0; Mismatches 55; Indels 77; Gaps 5;

QY 254 AGTTGATCCACTTGAACAGAGCAAAAGTGATTTGGTTCTGATACATTAATTCAA 313
DB 21 AGTTGATCCACTTGAACAGAGCAAAAGTGATTTGGTTCTGATACATTAATTCAA 80
QY 314 TGTTTTGGTTTATTTGGCAACCAAGAGTTATCTTAAGAACATCCAGTAAACAG 373
DB 81 TGTTTTGGTTTATTTGGCAACCAAGAGTTATCTTAAGAACATCCAGTAAACAG 140
QY 374 AATTGGAAGATCAGATATATATGAAACAGAGTCAAGAAATPAACAGAGAAAAAG 433
DB 141 AATTGGAAGATCAGATATATATGAAACAGAGTCAAGAAATPAACAGAGAAAAAG 200
QY 434 CTGGCAAGCTGACAGAGGTCAGCTTCAAGATTTGTAATAAATGCCCTGGGAACAA 493
DB 201 CTGGCAAGCTGACAGAGGTCAGCTTCAAGATTTGTAATAAATGCCCTGGGAACAA 260
QY 494 AATCGAAAAATGATCAAAAAGTGGCCAAATAAGAAAAAGTTAACTTTTGGTT 553
DB 261 AATCGAAAAATGATCAAAAAGTGGCCAAATAAGAAAAAGTTAACTTTTGGTT 320
QY 554 TTGATGACACATATTCAAAAAGTACATTAATATGTAATCAC----- 595
DB 321 TTGATGACACATATTCAAAAAGTACATTTTTCCTCCCTCCAGAAATTAATCTG 380
QY 596 -----AGTATA 602
DB 381 TGGCAGGACAGAGTTAAATGTAATTTCTTAATGAATTAATGTAATTAATTAAGTAATA 440
QY 603 TGTAAAGCTAAATAC--TTCTCTCCAAAGATCAATATCTTATTTGATTAACAGTACAGAT 661
DB 441 TGTAAAGCTAAATACCTTTCTTCCAAAGATGTTATCTTTATTTGATTTGACATGAGAT 500
QY 662 TTTAATGTTGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 719
DB 501 TTTAATGTTGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 560
QY 720 TTTCTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 779
DB 561 TTTCTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 620
QY 780 AATGTTCTTTGTAACATTTGTAATTTTAAATGAATATGACCTTATGATGATGATGAT 839
DB 621 AATGTTCTTTGTAACATTTGTAATTTTAAATGAATATGACCTTATGATGATGATGATGAT 680
```





**JOURNAL  
COMMENT**

Unpublished (1999)  
Contact: Robert Straube, Ph. D.  
Email: CGabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://www.imgc.org>

## FEATURES

## Bouyce

## ORIGIN

Query Match	51.6%	Score 596.6;	DB 4;	Length 1014;
Best Local Similarity	89.2%	Pred. No. 7.9e-15;		
Matches 725; Conservative	0;	Mismatches 21;	Indels 67;	Gaps 5

Qy	91	GGAGGGGTGAGGAGCAGAGCTGGCCATAATGGCAGGTGAAGAAATTAAGACATATCA	150
Db	33	GGAGTCGAAGGCCGTGAGGTCAAGCCATATGCGAGGTGAAGAAATTAAGACATATCA	92
Qy	151	GTGAATTTCA CGAGTATTTGTCAAGCTTTGAGATTCCATTGGTGTGTGATGAGATG	210
Db	93	GTAAATAATTCACGAGTATTTGTCAAGCGTTTGAGAAATTCATTGGTGTGTGATGAGATG	152
Qy	211	CTGAAGACCATGATGTCTGTTCTAGAATGAGTTGTGCAGAAATTTGGATCACTTGA	270
Db	153	CTGAAGACCATGATGTCTGTTCTAGAATGAGTTGTGCAGAAATTTGGATCACTTGA	212
Qy	271	CAAGCAAAAGTGGATTTGGTTTCTGCATACACATTAATCAATGTTTGGTTTATTTG	330
Db	213	CAAGCAAAAGTGGATTTGGTTTCTGCATACACATTAATCAATGTTTGGTTTATTTG	272
Qy	331	GCAACCCAGGAGTTAATCCTAAGGAACATCCAGTAAACAGAAATTTGGAAGAATTCGA	390
Db	273	GCAACCCAGGAGTTAATCCTAAGGAACATCCAGTAAACAGAAATTTGGAAGAATTCGA	332
Qy	391	GTATATATGAAACAGAGTCAGAGAAATTAACAGCAAGAAATAAGGCTGGCAAGCTGCAGCA	450
Db	333	GTATATATGAAACAGAGTCAGAGAAATTAACAGCAAGAAATAAGGCTGGCAAGCTGCAGCA	392
Qy	451	GGTCACAGCTCAAGATTTGTAAAAAATGCGCTGTGGGAACCAAAATTCGAAAAATGCATCA	510
Db	393	GGTCACAGCTCAAGATTTGTAAAAAATGCGCTGTGGGAACCAAAATTCGAAAAATGCATCA	452
Qy	511	AAAGTTCGCAATAAAGAAAAAGTAAAGTTAACTTTTGGTTTGTATGTACACATATTC	570
Db	453	AAAGTTCGCAATAAAGAAAAAGTAAAGTTAACTTTTGGTTTGTATGTACACATATTC	512
Qy	571	AAAAAGTAC-----	579
Db	513	AAAAAGTACCTTCCGCCACCCCCCGCAAAATTAATCTGTGGCAGGCGAAGTTTANA	572
Qy	580	-----ATTAAATGTATGATACA---GTAAATGTAAAGCTAAATAC-TTCCTCTC	625
Db	573	TGTGTTCTTATATATATGTAAATTCACAGTAATATGTAAAGCTAAATACCTTCTCTC	632
Qy	626	CAAGATCATATCTTATTTATTTAGCTGAGCACTGAGATTTAACTTGTGATATATATATA	685

Db	633	CAAGATCATATCTTTATTGATTAGACCTGAGCATTTTAACATTGNGATTAATTATATA	692
Qy	686	TTTATTAATTTAACCATCTCTGTATGAGACTCTTATTTCTTTATATAGGTGACGTCTTGCAAG	745
Db	693	TTTATTAATTTTACCATCTCTGTATGAGACTCTTATTTCTTTATATAGGTGACGTCTTGCAAG	752
Qy	746	TACCATTTTATTAACACGCTGTGAAATTTTAAGTGAATGTCTTTGTAAACATTTTGTACTA	805
Db	753	TACCATTTTATTAACACGCTGTGAAATTTTAAGTGAATGTCTTTGTAAACATTTTGTACTA	812
Qy	806	-TTTAAATGAATTAATGA-CCTTTATGAGATGAT	836
Db	813	TTTTTAAATGAATTAATGACCTTTATGAAATGATG	845

RESULT 6  
BG189347  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BG189347	847 bp	mRNA	linear	EST 21-APR-200
BS88391	Athersys	RAGE Library	Homo sapiens cDNA,	mRNA sequence.
BG189347				
BG189347.1	GI:13711034			
EST.				
Homo sapiens	(human)			
Homo sapiens				

## FEATURES

Location/Qualifiers  
1..847

## ORIGIN

511 AAAGTCCCAATAAAGGAAAAAGTAAAGTAACTTTTGGTTTGATGTACACATATTC 570  
|||||  
453 AAAGTCCCAATAAAGGAAAAAGTAAAGTAACTTTTGGTTTGATGTACACATATTC 512

Qy	279	AGGGAATTGGTTCTGCAATACATTAATTCATGTTTGGGTTTATTTGGCAACCA	338
Db	25	AGGGAATTGATTTCTGCAATACATTAATTCATGTTTGGGTTTATTTGGCAACCA	84
Qy	339	AGGATTAATCTTAAGAACATCCAGTAAACAGAAATTGGAAAGATCAGATATATAT	398
Db	85	AGGATTAATCTTAAGAACATCCAGTAAACAGAAATTGGAAAGATCAGATATATAT	144
Qy	399	GAACACAGTCAAGAAATTACACACAGAAAAAGGCTGGCAAGTGCACAGGTGCAGC	458

```

Db      145 GAACAGAGTCAGAGAAATAACAGACAGAAAAAGCTGGCAAGCTGGACAGAGGTGACG 204
Qy      459 TTCAAGATTGTGTAATAAATGCTCTGGGAA-ACCAAAATCGAAAAATGCATCAAAAGTTG 517
Db      205 TTCAAGATTGTGTAATAAATGCTCTGGGAAATCAAAATCGAAAAATGCATCAAAAGTTG 264
Qy      518 CCAATTAAGGAAAAAGTAAAGTAACTTTGGTGGTGGTACACATATTCAAAAGT 577
Db      265 CCAATTAAGGAAAAAGTAAAGTAACTTTGGTGGTGGTACACATATTCAAAAGT 324
Qy      578 AC----- 579
Db      325 ACATCCCCCCCCCCCCCGCAAAATATCTGGGCAAGGCAAGTTAAATGTC 384
Qy      580 -----ATTATATGTAATCACA---GTATATGTAAGCTAAATAC-TTCCTTCGAAA 629
Db      385 TTCTTATTAATATGTAATTAATCAGAGTAATATGTAAGCTAAATATCTTCCTCCAAA 444
Qy      630 GATCATTTATCTTATTTGTTAGCTAGAGATTTTAACTTTGATGATATATATATTTA 689
Db      445 GATCATTTATCTTATTTGTTAGCTAGAGATTTTAACTTTGATGATATATATATTTA 504
Qy      690 TAATTTACCATCTCTTGATGAGACTCTTAATTTCTTTATATAGTCAAGTCTGCAAGTACC 749
Db      505 TAAATTCACATCTCTTGATGAGACTCTTAATTTCTTTATATAGTCAAGTCTGCAAGTACC 564
Qy      750 ATTTTATAGAGAGCTGGAATTTAAGTGAATGTTCTTTGTAACATTTGTAATTTT 809
Db      565 ATTTTATAGAGAGCTGGAATTTAAGTGAATGTTCTTTGTAACATTTGTAATTTT 624
Qy      810 AAATGAATTAAGACTTATGATGATGCTGCTGAGGCTGAATTTATGTAATCTGT 869
Db      625 AAATGAATTAAGACTTATGATGATGCTGCTGAGGCTGAATTTATGTAATCTGT 684
Qy      870 TTTCACATATGATTAATTAAGAGCGTG-AAATGCTTAATGTTGATTTTCTGATA 928
Db      685 TTTCACATATGATTAATTAAGAGCGTGAAAGCTGAATGCTAAATGTTCTGATA 744
Qy      929 GATATCTTATGATTTTATGATGATTTAGAAATGCTGTTGTTGTAATTCAAAGTGTG 988
Db      745 GATATCTTATGATTTTATGATGATTTAGAG-ATTAATGCTGTTGTTGTAATTCAAAGTGTG 803
Qy      989 AAATGAATTTAGTGTGTTGATCTTAATTTCTGGCATGTTGC 1031
Db      804 AAATGAATTTAGTGTGTTGATCTTAATTTCTGGCATGTTGC 846

```

```

RESULT 7
LOCUS   BG217959          853 bp      mRNA      linear      EST 21-Apr-2001
DEFINITION R537683 Atherysa RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG217959
VERSION   BG217959.1 GI:13743980
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS  Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
          Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Booser,S.,
          Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S.,
          Mays,R., Smith,E., Veloso,N., Klika,A., Hesse,J., Cochren,K., Lo,K.,
          Offenbacher,J., Danzig,J. and Ducar,M.,
          Creation of genome-wide protein expression libraries using random
          activation of gene expression
          Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL  MEDLINE
PUBMED   11329013
COMMENT   Contact: Scott J. Cain
          Athersys, Inc.
          3201 Carnegie Ave, Cleveland, OH 44115, USA
          Tel: 216 431 9900

```

```

FEATURES
Source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HT1080"
/clone_id="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

```

```

Query Match      51.1%; Score 590.2; DB 4; Length 853;
Best Local Similarity 87.5%; Pred. No. 1,7e-113;
Matches 743; Conservative 0; Mismatches 33; Indels 73; Gaps 6;

```

```

Qy      323 TTTATTTGCAACCCAGAGATTAACTCTAGAAATCATCAGTAACAGGAATTGAAA 382
Db      4 TTTATTTGCGCACCCAGAGATTAACTCTAGAAATCATCAGTAACAGGAATTGAAA 63
Qy      383 GAATCAGAGATATATGAAACAGAGTCAGAGAAATTAACAGCAAGAAAAAGCTGGCAAGC 442
Db      64 GAATCAGAGATATATGAAACAGAGTCAGAGAAATTAACAGCAAGAAAAAGCTGGCAAGC 123
Qy      443 TGGACAGAGCTGACGCTTCAAGATTGTAATAAATGCCCTCTGGGAAACCAAAATCGAAA 502
Db      124 TGGACAGAGCTGACGCTTCAAGATTGTAATAAATGCCCTCTGGGAAACCAAAATCGAAA 183
Qy      503 ATGCATCAAAAGTTGCCAATAAAGGAAAAAGTAAAGTAACTTTGGTTGATGTAC 562
Db      184 ATGCATCAAAAGTTGCCAATAAAGGAAAAAGTAAAGTAACTTTGGTTGATGTAC 243
Qy      563 ACATATTCAAAAGTAC----- 579
Db      244 ACAATATCAAAAAGTACATCTTCCCCCCCCCCCCCGCAAAATATCTGTGGCAGG 303
Qy      580 -----ATTATATGATATACA---GTATATGTAAGCTAAA 614
Db      304 GCAAGTTTAATGTTGTTCTTATTAATATGTAATTAATCAAGTAATATGTAAGCTAAA 363
Qy      615 TAC-TTCTCTCCAAAGATATATCTTTATGATTAGCACTGAGATTAAACATGTGG 673
Db      364 TACTTCTCTCCAAAGATATATCTTTATGATTAGCACTGAGATTAAACATGTGG 423
Qy      674 ATATATATATATTTAATTTACCATCTCTGATGAGACTCTTAATTTCTTATATAGT 733
Db      424 ATATATATATATTTAATTTACCATCTCTGATGAGACTCTTAATTTCTTATATAGT 483
Qy      734 CAGCTTGGCAAGTACATTTTATAGACAGCTGTGAATTTAAGTAAGTCTTTGTA 793
Db      484 CAGCTTGGCAAGTACATTTTATAGACAGCTGTGAATTTAAGTAAGTCTTTGTA 543
Qy      794 ACATTTGTAATTTTAAATGTAATAGACCTTATGAAAGTATGCTATCTGAGCTGAAA 853
Db      544 ACATTTGTAATTTTAAATGTAATAGACCTTATGAAAGTATGCTATCTGAGCTGAAA 603
Qy      854 TTTATAGTATCATCTGTTTCACTATATGATATTAAGAAAGCGGAAATGACTTAAATGTT 912
Db      604 TTTATAGTATCATCTGTTTCACTATATGATATTAAGAAAGCGGAAATGACTTAAATGTT 663
Qy      913 CATTTTTCGTATAGATCTTATATCATGTTTCAATGATTTTGAAGATTA-CTGCTTGG 971
Db      664 CATTTTTCGTATAGATCTTATATCATGTTTCAATGATTTTGAAGATTAACCTGCTTAG 723
Qy      972 TTGATATTCAAAGTGTGAACTAAAGTTATGTTGTAC-TTTAATCTTGGCATGTT 1029
Db      724 CTGAAATTCAAAGTGTGAACTAAACCTTAAGGTTGACCTTAAATTAATCTGGCATGTT 783

```

QY 1030 GCCTCTATGTCCTATTAATAATACATTCCTTAACCTTGTAGTGGAAATAAGT 1089  
| | | | |  
Db 784 GCCCTTAATGTCCTATTAATAATACATTCCTTAACCTTGTAGTGGAAATAAGT 843  
| | | | |  
QY 1090 TGTATGTTG 1098  
| | | | |  
Db 844 TTGATGTTG 852  
| | | | |  
RESULT 8  
BG185663 849 bp mRNA linear EST 21-APR-2001  
LOCUS RST4614 Atherys RAGE Library Homo sapiens cDNA, mRNA sequence.  
DEFINITION BG185663  
ACCESSION BG185663  
VERSION BG185663.1 GI:13707350  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 849)  
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,  
Cain,S., Leventhal,C., Thornton,M., Ramchandran,R.,  
Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S.,  
Mayer,R., Smith,E., Veloso,N., Kika,A., Hess,J., Cochren,K., Lo,K.,  
Offenbacher,J., Danzig,J. and Ducar,M.  
Creation of genome-wide protein expression libraries using random  
activation of gene expression  
Nat. Biotechnol. 19 (5), 440-445 (2001)  
21227151  
JOURNAL MEDLINE  
PUBMED 11329013  
COMMENT Contact: Scott J. Cain  
Atherys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: scain@atherys.com  
High quality sequence stop: 505.  
Location/Qualifiers  
1. 849  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/cell\_line="HT1080"  
/note="See 'Atherys RAGE Library'  
Libraries using Random Activation of Gene Expression',  
Nature Biotechnology, in press. Note that even though the  
cell type indicated is HT1080, since a random activation  
method was used, these sequence tags are not necessarily  
expressed in HT1080 under normal circumstances."  
ORIGIN  
Query Match 50.64; Score 584.6; DB 4; Length 849;  
Best Local Similarity 89.34; Pred. No. 2.6e-112;  
Matches 723; Conservative 0; Mismatches 20; Indels 67; Gaps 6;  
QY 263 CACTTGAACAGCAAAAGTGGATTTGTTCTGATACATTAATTCATGTTTGGG 322  
| | | | |  
Db 10 CCCTTAACAAGCAAAAGTGGATTTGTTCTGATACATTAATTCATGTTTGGG 69  
| | | | |  
QY 323 TTTATTTGGCAACCCAGAGTTTATCTTAAGAAACATCCAGTAAACAGAAATTGAAA 382  
| | | | |  
Db 70 TTTATTTGGCAACCCAGAGTTTATCTTAAGAAACATCCAGTAAACAGAAATTGAAA 129  
| | | | |  
QY 383 GAATCAGATATATGAACAAGTCAAGAAATTAACAACAAGAAAGGCTGGCAAGC 442  
| | | | |  
Db 130 GAATCAGATATATGAACAAGTCAAGAAATTAACAACAAGAAAGGCTGGCAAGC 189  
| | | | |  
QY 443 TGGACAGAGGTGACGTTCAAGATTTGTAAATAATGCCCTCTGGGAACCAAAATCGAAA 502  
| | | | |  
Db 190 TGGACAGAGGTGACGTTCAAGATTTGTAAATAATGCCCTCTGGGAACCAAAATCGAAA 249  
| | | | |

QY 503 ATGCATCAAAAGTGGCAATTAAGAAAGAAAGTAAAGTAACTTTTGTGTTGATGAC 562  
| | | | |  
Db 250 ATGCATCAAAAGTGGCAATTAAGAAAGAAAGTAAAGTAACTTTTGTGTTGATGAC 309  
| | | | |  
QY 563 ACATATTCAAAAGTAC----- 579  
| | | | |  
Db 310 ACATATTCAAAAGTAC----- 369  
| | | | |  
QY 580 -----ATTAATATGATACAC---GTAAATGTAAAGCTAAATTC- 618  
| | | | |  
Db 370 GTTAAATGTTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTC 429  
| | | | |  
QY 619 TCCCTCCAAAGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 678  
| | | | |  
Db 430 TCCCTCCAAAGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 489  
| | | | |  
QY 679 TTATATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 738  
| | | | |  
Db 490 TTATATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 549  
| | | | |  
QY 739 TTGCAAGTACATTTTATTAAGACGCTGTGAATTTAAGTGAATGTTCTTTGTAACATT 798  
| | | | |  
Db 550 TTGCAAGTACATTTTATTAAGACGCTGTGAATTTAAGTGAATGTTCTTTGTAACATT 609  
| | | | |  
QY 799 TGTACTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 858  
| | | | |  
Db 610 TGTACTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 669  
| | | | |  
QY 859 GTTACATCTGTTTTCACCTA-TATGATATTAAGAAAGCTG-AATGACTTAAATGTTCAAT 916  
| | | | |  
Db 670 GTTACATCTGTTTTCACCTA-TATGATATTAAGAAAGCTG-AATGACTTAAATGTTCAAT 729  
| | | | |  
QY 917 TTTTCTGATAGATACCTTATGATGTTTCAATGATTTTGAATTAATGCTTTGTTGAT 976  
| | | | |  
Db 730 TTTTCTGATAGATACCTTATGATGTTTCAATGATTTTGAATTAATGCTTTGTTGAT 789  
| | | | |  
QY 977 ATTCAAGTCTGAACCTTAAAGTATTATGCT 1006  
| | | | |  
Db 790 A-TCAAGGCTGAACCTTAAAGTATTATGCT 818  
| | | | |  
RESULT 9  
B1753996 878 bp mRNA linear EST 25-SEP-2001  
LOCUS B1753996  
DEFINITION 603027431F1 NIH\_MGC\_114 Homo sapiens cDNA clone IMAGE:519799 5',  
mRNA sequence.  
ACCESSION B1753996  
VERSION B1753996.1 GI:15745574  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 878)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
http://image.jnl.gov  
Plate: LIAH1495 row: m column: 08  
High quality sequence stop: 843.  
Location/Qualifiers  
1. 878  
/organism="Homo sapiens"  
/mol\_type="mRNA"



QY 433 GCTGGAGCTGAGCAGAGGTGAGCTTCAAGATTGTAATAAATGCTCTGGAAACA 492  
 DB 243 GCTGGAGCTGAGCAGAGGTGAGCTTCAAGATTGTAATAAATGCTCTGGAAACA 302  
 QY 493 AATCGAAAAATGATCAAAAAGTCCCAATTAAGAAAAAGTAAAGTAACTTTGGT 552  
 DB 303 AATCGAAAAATGATCAAAAAGTCCCAATTAAGAAAAAGTAAAGTAACTTTGGT 362  
 QY 553 TTGATGTACATATTTCAAAAAGTAC-----579  
 DB 363 TTGATGTACATATTTCAAAAAGTACATTTCCCCCCCCCCCCCGCAAAATAT 422  
 QY 580 -----ATTAAATATGTA-----TCAAGTATAT 603  
 DB 423 TCTGTGAGGAGGCAAGCTTTAAATGCTTTCTTATTAATGTAATTTACAGTAATAT 482  
 QY 604 GTAAGCTAAATAC-TTCTCTCCCAAGATCATTTATTTGATTAAGCACTGAGATT 662  
 DB 483 GTAAGCTAAATACCTTCTCTCCCAAGATCATTTATTTGATTAAGCACTGAGATT 542  
 QY 663 TTAACATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 722  
 DB 543 TTAACATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 602  
 QY 723 TTAATATAGGCTCTGCAAGTATTAATTAATTAATTAATTAATTAATTAATTAAT 782  
 DB 603 TTAATATAGGCTCTGCAAGTATTAATTAATTAATTAATTAATTAATTAATTAAT 662  
 QY 783 GTTCTTTGTAACATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 842  
 DB 663 GTTCTTTGTAACATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 722  
 QY 843 GTTGGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 892  
 DB 723 GTTGGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 772

RESULT 11  
 AL540005/c 887 bp mRNA linear EST 24-MAR-2004  
 LOCUS AL540005 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
 DEFINITION CS0DF031YH10 3-PRIME, mRNA sequence.  
 ACCESSION AL540005  
 VERSION AL540005 GI:45715641  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 15, 2001 this sequence version replaced gi:31264568.  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: sequef@genoscope.cns.fr, web: www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library  
 was not normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.  
 This sequence belongs to sequence cluster 1461.r  
 For more information about this cluster, see  
 http://www.genoscope.cns.fr/cdna?c=CS0DF031YH10&c=1461.r.

FEATURES  
 source  
 1..887  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF031YH10"

/issue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_lib="Homo sapiens FETAL BRAIN"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with Not I and  
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."

ORIGIN

Query Match 49.8%; Score 575.4; DB 1; Length 887;  
 Best Local Similarity 78.4%; Pred. No. 2.2e-110;  
 Matches 695; Conservative 65; Mismatches 51; Indels 75; Gaps 5;  
 QY 324 TTAATTTGCAACCCAGAGATTAATCTTAAGAAATCCAGTAATTAACAGAAATGGAAAG 383  
 DB 886 TTTTATTTGCAACCCAGAGATTAATCTTAAGAAATCCAGTAATTAACAGAAATGGAAAG 827  
 QY 384 AATCAGATTAATTAAGACAGAGTCAAGAAATTAACAGACAGAAAGGCTGGCAAGCT 443  
 DB 826 AATCAGATTAATTAAGACAGAGTCAAGAAATTAACAGACAGAAAGGCTGGCAAGCT 767  
 QY 444 GACAGAGTGAAGCTTCAAGATTTGTAATAAATGCTCTGGCAACCAAAATCGAAAAA 503  
 DB 766 GGRMRGGGGGSGCTTCAAGATTTTAAAAAAMDCCTTGGGACCAAAATGAAAAA 707  
 QY 504 TGATCAAAAAGTTGCCAATTAAGAAAAAGTAAAGTAACTTTTGGTTTGAATGACA 563  
 DB 706 AGVAAAAAARTRGCCAAAAAAGAAAAAAMAAAAATTAATTTTGGTTTGAATGAAAA 647  
 QY 564 CATATCAAAAAGTAC-----579  
 DB 646 AAWATTAATAAATAATTAATTTTCCCCCCCCCCCCCGCACAAATTAATCTGGGCA 587  
 QY 580 -----ATTAAATATGTA-----TCAAGTATAT 603  
 DB 586 GGGCAAGCTTAATGATGCTCCCTTATTAATTAATTAATTAATTAATTAATTAATTAAT 527  
 QY 613 AATACCTT-CTCTCCCAAGATCATTTATTTGATTAAGCACTGAGATTTTAACTTG 671  
 DB 526 AATACCTTCCCCCCCCCAAGATCATTTATTTGATTAAGCACTGAGATTTTAACTTG 467  
 QY 672 TGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 731  
 DB 466 TGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 407  
 QY 732 GTGAGCTTGAAGTACATTTTAAAGAGAGCTGTAATTAAGTAAAGTCTTTTGT 791  
 DB 406 GTGAGCTTGAAGTACATTTTAAAGAGAGCTGTAATTAAGTAAAGTCTTTTGT 347  
 QY 792 AAAATTTGATTAATTTAAATGAATTAAGCTTAATGAAGTATGCTATCTGAGCTGA 851  
 DB 346 AAAATTTGATTAATTTAAATGAATTAAGCTTAATGAAGTATGCTATCTGAGCTGA 287  
 QY 852 AATTATAGTAACTCTGTTTCACTATGATTAATTAAGAAAGGCG-AATGAATTAATG 910  
 DB 286 AATTATAGTAACTCTGTTTCACTATGATTAATTAAGAAAGGCG-AATGAATTAATG 227  
 QY 911 TTCAATTTTTCGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 970  
 DB 226 GTCATTTTTCGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 167  
 QY 971 GTTGATTAATTAAGTGAAC--TAAAGTTATGTTGATTTAATTTTGGCATGT 1028  
 DB 166 GTTGATTAATTAAGTGAACCTCAAACTKTRTGTTGATTTAATTTCTGGCATGT 107  
 QY 1029 TGGCTTATGCTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1088  
 DB 106 TGGCTTATGCTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 47  
 QY 1089 TTGTATGTTGATGAATTAATTTGGCATGATGACTGACTGCAAT 1134  
 DB 46 TTGTATGTTGATGAATTAATTTGGCATGATGACTGACTGCAAT 1

RESULT 12  
 LOCUS BM679939/c  
 DEFINITION UI-E-BOI-aiX-a-14-0-UI-81 UI-E-BOI Homo sapiens linear EST 27-FEB-2002  
 ACCSSION UI-E-BOI-aiX-a-14-0-UI 3', mRNA sequence.  
 VERSION BM679939  
 KEYWORDS EST.  
 SOURCE GI:18989835  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 807)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PUBMED 8889548  
 COMMENT Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@iowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing Arrayed by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 The following repetitive elements were found in this cDNA sequence: 476-498, >AT-richlow\_complexity (matched complement)  
 Seq primer: M13 Forward  
 POLYA=yes.

FEATURES  
 source  
 location/Qualifiers  
 1..807  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-BOI-aiX-a-14-0-UI"  
 /tissue\_type="fetal eye"  
 /dev\_stage="fetal eye"  
 /lab\_host="NH108 (Life Technologies) (T1 phage resistant)"  
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-BOI is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag (dTT)18 tail. The sequence tag for this library is CGCGTATACC. This library was created for the program Gene Discovery in the Visual System, supported by National Eye Institute (NEI).  
 TAG TISSUE=human fetal eye  
 TAG LIB=UI-E-BOI  
 TAG\_SEQ=CGCGTATACC"

ORIGIN  
 Query Match 49.1%; Score 568; DB 4: Length 807;  
 Best Local Similarity 89.8%; Pred. No. 7.9e-109;  
 Matches 692; Conservative 0; Mismatches 11; Indels 68; Gaps 5;

QY 452 GTGAGCTTCAAGATTGTAAAAATGCCCTTGGAACCAAAATCGAAAAATCATCAA 511  
 DB 783 GTGAGCTTCAAGATT--GTAAAAATGCCCTTGGAACCAAAATCGAAAAATCATCAA 726  
 QY 512 AAGTGGCAATTAAGAAAAAGTAAAGTAACTTTTGGTTTGATGATACATATTTCA 571  
 DB 725 AAGTGGCAATTAAGAAAAAGTAAAGTAACTTTTGGTTTGATGATACATATTTCA 666  
 QY 572 AAAAGTAC----- 579  
 DB 665 AAAAGTACATCTCCGCCACCCGCCAAAAATTTGTGGACAGGCAAGTTAAAT 606  
 QY 580 -----ATTATATGATATACA--GTAAATGTAAGCTAAATAC--TTCCCTCC 626  
 DB 605 GTGTTCTTATATATATGTAATTCACAGTAATATGTAAGCTAAATATCTTCTCC 546  
 QY 627 AAAGTCAATTCCTTTATTTGATTTAGCAGTGAAGTTTAACTTTGATATATATAT 686  
 DB 545 AAAGTCAATTCCTTTATTTGATTTAGCAGTGAAGTTTAACTTTGATATATATAT 486  
 QY 687 TTATATTTACATCTCTTGATGAGACCTTTATTTCTTATATAGTCACTTGAAGT 746  
 DB 485 TTATATTTACATCTCTTGATGAGACCTTTATTTCTTATATAGTCACTTGAAGT 426  
 QY 747 ACCATTTTAAAGCAGCTGTGAATTTAAGTGAATGTCTTTGTAACATTTGACTAT 806  
 DB 425 ACCATTTTAAAGCAGCTGTGAATTTAAGTGAATGTCTTTGTAACATTTGACTAT 366  
 QY 807 TTTAAATGAATATACCTTTTGAAGTATCTTATGAGCTGTAATATAGTATATC 866  
 DB 365 TTTAAATGAATATACCTTTTGAAGTATCTTATGAGCTGTAATATAGTATATC 306  
 QY 867 TGTTCCTATATGATATTTAAGAAAGCGT--AATGACTTAATATGCTATTTTCTGT 925  
 DB 305 TGTTCCTATATGATATTTAAGAAAGCGT--AATGACTTAATATGCTATTTTCTGT 246  
 QY 926 ATGATATCTTATATGATATTTTCAATTTAGAAATTTAGCTTTGATATTTCAAGT 985  
 DB 245 ATGATATCTTATATGATATTTTCAATTTAGAAATTTAGCTTTGATATTTCAAGT 186  
 QY 986 GTGAATCAATTAAGTTTATGTTGATATTTTATTTTATTTTATTTTATTTTATTT 1045  
 DB 185 GTGAATCAATTAAGTTTATGTTGATATTTTATTTTATTTTATTTTATTTTATTT 126  
 QY 1046 TAAATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1105  
 DB 125 TAAATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 66  
 QY 1106 AATTTGGCATGATGATCTGTCTCTCAATTAAGCTGAAATGTTGTAA 1156  
 DB 65 AATTTGGCATGATGATCTGTCTCTCAATTAAGCTGAAATGTTGTAA 15

RESULT 13  
 LOCUS AI800532/c  
 DEFINITION tc12a11.x1 Soares NhHMPu S1 Homo sapiens cDNA clone IMAGE:2063612  
 ACCSSION AI800532  
 VERSION AI800532  
 KEYWORDS AI800532.1 GI:5366004  
 SOURCE EST.  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 909)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov



This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1095 Std. Error: 0.00

Seq primer: -40UP from Gldco

High quality sequence stop: 407.

Location/Qualifiers

1. 909

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:2063612"

/tissue\_type="Pooled human melanocyte, fetal heart, and pregnant uterus"

/lab\_host="DH10B"

/clone\_1lb="Soares Nbh199.S1"

/note="Organ: mixed (see below); Vector: pTTT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBH, pregnant uterus Nbh199, and fetal heart Nbh199) were mixed, and six circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M.A.G.B. clones 260232-265223, 340488-345479, and 484488-489479."

# ORIGIN

Query Match 49.0%; Score 567; DB 1; Length 909;  
Best Local Similarity 84.3%; Pred. No. 1.3e-108;  
Matches 759; Conservative 0; Mismatches 72; Indels 69; Gaps 8;

```

QY 325 TATTGGCAACCCAGAGATTATCTTAAGACATCTCAATAACAGCAATTGGAAGA 384
DB 909 TTTTGGCCACACGAGGTTAACTTAAGACCAT-CAGTTAAACCGGATTGGAAAG 851
QY 385 ATCAGATATATATGAACAGGTGCAAGAAATACACAAAGAAAGGCT-GGCAAGCT 443
DB 850 AATCAAGGTTTATTGACCAAGTCAAGAAATACCAAGCAAAAAGGCTGGCAAGTT 791
QY 444 GGAAGAGTGCAGCTTCAAGA-TTTGTAAAAATGCTCTGGGAACCAAAATCGAAAA 502
DB 790 GGCAGAGTGCAGGCTTCAGATTGTTAAAAATGGCTTTGGGACCCAAACCCGAAAA 731
QY 503 ATGCATCAAAAGTCCCAATTAAGAAAAAGTAAAGTTAATCTTTGGTTTGATGTAC 562
DB 730 ATGCATCAAAAGTCCCAATTAAGAAAAAGTAAAGTTAATCTTTGGTTTGATGTAA 671
QY 563 ACATATTCAAAAAGTAC-----579
DB 670 CAATTCAAAAGGTACATCCCCCCCCCCCCCAAAATTAATCTGTGGCAGGCGAAG 611
QY 580 -----ATTAATATGTAATACAC---GTAATATGTAAGCTAAATAC-T 618
DB 610 GTTAAATATGTTTCTTATTAATATGTAATATGCAATTAATATGTAAGCTAAATCTT 551
QY 619 TCCCTCCCAAGATCATTAATCTTAATGATTAAGCACTGAGGATTTTAACATTTGATATA 678
DB 550 TCCCTCCCAAGATCATTAATCTTAATGATTAAGCACTGAGGATTTTAACANTGATATA 491
QY 679 TTAATATTAATTAATACATCTTGATGAGACTCTTAATTTCTTAATATAGTGTGTC 738
DB 490 -TAAATATTAATTAATTAATCAATCTTGATGAGACTCTTAATTTCTTAATATAGTGTGTC 432
QY 739 TTCAAGTACATTTTATTAAGAGCTGTGAATTTTAAGTAAATGTTCTTTGTAACATT 798
DB 431 TTCAAGTACATTTTATTAAGAGCTGTGAATTTTAAGTAAATGTTCTTTGTAACATT 372
QY 799 TGAATATTTTAAATGAATATAGACCTTAATGAGATGCTATCTGTAAGGCTGAATATA 858
DB 371 TGAATATTTTAAATGAATATAGACCTTAATGAGATGCTATCTGTAAGGCTGAATATA 312
QY 859 GGAATATCTGTTTTCATATATGATATTAAGAAAGCGTG-AATGACTTAATATGTTCAATT 917

```

```

DB 311 GGAATATCTGTTTTCATATATGATATTAAGAAAGCGTGAAGTAAATGTTCAATT 252
QY 918 TTTTCTGTAATGATCTTATTAATGATTTTTCATGATTTTAAAGAAATTAATGTTGATA 977
DB 251 TTTTCTGTAATGATCTTATTAATGATTTTTCATGATTTTAAAGAAATTAATGTTGATA 192
QY 978 TTCAAGTACATTTTATTAAGAGCTGTGAATTTTAAGTAAATGTTCTTTGTAACATT 1037
DB 191 TTCAAGTACATTTTATTAAGAGCTGTGAATTTTAAAGTAAATGTTCTTTGTAACATT 132
QY 1038 GTCCATTTTAAATTAATTAATGATCTTCAATTAATTAATGATGAGAAATTAAGTGTATGTT 1097
DB 131 GTCCATTTTAAATTAATTAATGATCTTCAATTAATTAATGATGAGAAATTAAGTGTATGTT 72
QY 1098 GATGATGAATTTTGGCATGATGATGATCTTCAATTAATTAAGGCTGAATATGTTGATA 1156
DB 71 GATGATGAATTTTGGCATGATGATGATCTTCAATTAATTAAGGCTGAATATGTTGATA 13

```

RESULT 14  
BG213559  
LOCUS BG213559 1209 bp mRNA linear EST 21-APR-2001  
DEFINITION RST3166 Atherys RAGE Library Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG213559  
VERSION BG213559.1 GI:13735246  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1209)  
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,  
Cain,S., Leventhal,C., Thornton,M., Ramchandran,R.,  
Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S.,  
Mays,R., Smith,E., Veloso,N., Klike,A., Hess,J., Cochren,K., Lo,K.,  
Offenbacher,J., Danzig,J. and Ducar,M.  
Creation of genome-wide protein expression libraries using random  
activation of gene expression  
Nat. Biotechnol. 19 (5), 440-445 (2001)

JOURNAL MEDLINE 21227151  
PUBMED 11329013  
COMMENT Contact: Scott J. Cain  
Athersys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: scain@athersys.com

High quality sequence stop: 566.  
Location/Qualifiers  
1. 1209  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/cell\_line="HT1080"

/note="See 'Creation of Genome-wide Protein Expression  
Libraries using Random Activation of Gene Expression',  
Nature Biotechnology, in press. Note that even though the  
cell type indicated is HT1080, since a random activation  
method was used, these sequence tags are not necessarily  
expressed in HT1080 under normal circumstances."

## FEATURES

source

## ORIGIN

Query Match 48.9%; Score 564.8; DB 4; Length 1209;  
Best Local Similarity 86.1%; Pred. No. 3.8e-108;  
Matches 749; Conservative 0; Mismatches 46; Indels 75; Gaps 9;

```

QY 256 TTGATCACTTGAACAGCAAGAGTATTTGTTTTCATACATTAATTAATCAATG 315
DB 5 TTGAAACCACTGTAACAGCAAGAGTATTTGTTTTCATACATTAATTAATCAATG 63
QY 316 TTTTGGGTTTATTTTGGCAACCCAGAGATTATCTTAAGAAACATGCAATTAACAGAA 375

```

```

Db      64 TTTTGGGTTATTGGCAACCGAGAGTTAATCTTAAAGAAACATCTGTAAACAGANA 123
Qy      376 TTGGAAGAATCATGAGATATATATGAAACAGATCAAGAAATTAACAGANAAAAAGCT 435
Db      124 TTGGAAAGATCATGAGATATATATGAAACAGATCAAGAAATTAACAGANAAAAAGCT 183
Qy      436 GGCAAGCTGACAGAGAGTGCAGCTTCAAGATTTGTAATAATGCTCTTGGAAACAAA 495
Db      184 GGCAAGCTGACAGAGAGTGCAGCTTCAAGATTTGTAATAATGCTCTTGGAAACAAA 243
Qy      496 TCGAATAATGCTCAAAAGTTGCCAATAAGAAAAGTAAAGTTAACTTTTGGTTT 555
Db      244 TCGAATAATGCTCAAAAGTTGCCAATAAGAAAAGTAAAGTTAACTTTTGGTTT 303
Qy      556 GATGTACACATATTCAAAAAGTAC----- 579
Db      304 GATGTACACATATTCAAAAAGTACCTTCCCCCCCCCCCCCGCAAAATATCTGT 363
Qy      580 -----ATTAAATATGTAATCACA---GTAATATGTAA 608
Db      364 GCGAGGGCAAGTTTAAATGTGTTTCTTATTAATAGCTAAATTCACAGTAATATGTAA 423
Qy      609 GCTAAATAC-TTCCCTCCAAAGATCATATCTTTTATGTAAGCACTGAGATTTTAAC 667
Db      424 GCTAAATACCTTCTCCAAAGATCATATCTTTTATGTAAGCACTGAGATTTTAAT 483
Qy      668 ATTGTATATATATATATATATATATATATATATATATATATATATATATATAT 727
Db      484 ATTGTATATATATATATATATATATATATATATATATATATATATATATATAT 543
Qy      728 ATAGTCACTGTCAGATCACTTATTAAGAGCTGTGAATTAAGTGAATGTCT 787
Db      544 ATAGTCACTGTCAGATCACTTATTAAGAGCTGTGAATTAAGTGAATGTCT 603
Qy      788 TTGTAAACATTTTACTATTTTAAATGAATGAACCTTAAGATGATGATCTGATG 847
Db      604 TTGTAAACATTTTACTATTTTAAATGAATGAACCTTAAGATGATGATCTGATG 663
Qy      848 CTGAATATTAAGTACATCTGTTTCACTATGATATGAATGAATGAATGAATGATG 906
Db      664 CTGAATATTAAGTACATCTGTTTCACTATGATATGAATGAATGAATGAATGATG 723
Qy      907 AATGTCATTTTCTGATATGATATGATATGATATGATATGATATGATATGATATG 966
Db      724 AATGTCATTTTCTGATATGATATGATATGATATGATATGATATGATATGATATG 783
Qy      967 CTGTTGATATTCAGAGTGAACCTTAAGTGAATGATATGATATGATATGATATG 1026
Db      784 GCTTGTATGATATTCAGAGTGAACCTTAAGTGAATGATATGATATGATATGATATG 839
Qy      1027 GTTGCCTGTATGATTCATTTAAATTAAT 1056
Db      840 GTTG-CTCTATGTCATTAATTAATTAAT 868

```

RESULT 15  
 LOCUS BG206383 859 bp mRNA linear EST 21-APR-2001  
 DEFINITION R525831 Atherys RAGE Library Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BG206383  
 VERSION BG206383.1 GI:13728070  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 859)  
 Harrington, J. J., Sherf, B., Rundlett, S., Jackson, P. D., Perry, R.,  
 Cain, S., Leventhal, C., Thornton, M., Ramchandran, R.,  
 Whitlington, J., Lerner, L., Costanzo, D., McElligott, K., Booser, S.,  
 Mays, R., Smith, E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K.,  
 Offenbacher, J., Danzig, J., and Ducar, M.

TITLE Creation of genome-wide protein expression libraries using random  
 activation of gene expression  
 JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)  
 MEDLINE 21227151  
 PUBMED 11329013  
 COMMENT  
 Contact: Scott J. Cain  
 Atherys, Inc.  
 3201 Carnegie Ave, Cleveland, OH 44115, USA  
 Tel: 216 431 9900  
 Fax: 216 361 9596  
 Email: scain@atherys.com  
 High quality sequence stocp: 513.  
 Location/Qualifiers  
 1..859  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /cell\_line="HT1080"  
 /clone\_lib="Atherys RAGE Library"  
 /note="See 'Creation of Genome-wide Protein Expression  
 Libraries using Random Activation of Gene Expression',  
 Nature Biotechnology, in press. Note that even though the  
 cell type indicated is HT1080, since a random activation  
 method was used, these sequence tags are not necessarily  
 expressed in HT1080 under normal circumstances."

## ORIGIN

Query Match 48.7%; Score 562.4; DB 4; Length 859;  
 Best Local Similarity 88.1%; Pred. No. 1.2e-107;  
 Matches 754; Conservative 0; Mismatches 26; Indels 76; Gaps 10;

```

Qy      263 CACTGAACAGCAAAAGTGAATTTGTTCTGATACATTAATTAATGTTGG 322
Db      10 CCGTGAAACAGCAAAAGTGAATTTGTTCTGATACATTAATTAATGTTGG 69
Qy      323 TTTATTTGGCAACCAAGAGTTAACTTAAGAACTCCGTAATAACAGAAATGGA 382
Db      70 TTTATTTGGCAACCAAGAGTTAACTTAAGAACTCCGTAATAACAGAAATGGA 129
Qy      383 GAATCAGATATATATGAAACAGATCAAGAAATTAACAGAAAGGCTGGCAAGC 442
Db      130 GAATCAGATATATATGAAACAGATCAAGAAATTAACAGAAAGGCTGGCAAGC 189
Qy      443 TCGACAGAGTGCAGCTTCAAGATTTGTAATAATGCTCTGCGAAACCAAAATCGAAA 502
Db      190 TCGACAGAGTGCAGCTTCAAGATTTGTAATAATGCTCTGCGAAACCAAAATCGAAA 249
Qy      503 ATGCATCAAAAGTTCAGATTAAGAAAGTAAAGTTAACTTTTGGTTGATGTAC 562
Db      250 ATGCATCAAAAGTTCAGATTAAGAAAGTAAAGTTAACTTTTGGTTGATGTAC 309
Qy      563 ACATATTCAAAAGTAC----- 579
Db      310 ACATATTCAAAAGTACATTCCTCCCCCCCCCCCCCGCAAAATTAATTCGTGGCAGG 369
Qy      580 -----ATTAAATGTAATCACA---GTAATATGTAAGCTAAT 615
Db      370 CAAGGTTAAATGTTGTTCTTATTAATGTAATTAATTAATTAATTAATTAATTAAT 429
Qy      616 AC-TTCCCTCCAAAGATCATATCTTTTATGTAAGCACTGAGATTTTAACATTTGTA 674
Db      430 ACTTCCCTCCAAAGATCATATCTTTTATGTAAGCACTGAGATTTTAACATTTGTA 489
Qy      675 TATATATATATATATATATATATATATATATATATATATATATATATATATAGTC 734
Db      490 TATATATATATATATATATATATATATATATATATATATATATATATATATAGTC 549
Qy      735 AGCTTGCAGATCACTTTATAGCACTGTGAATTAAGTGAATTAAGTGAATTAAGTGA 794
Db      550 AGCTTGCAGATCACTTTATAGCACTGTGAATTAAGTGAATTAAGTGAATTAAGTGA 609
Qy      795 CATTTGACTATTTTAAATGAATATAGCTTAAGATATGCTATCTGAGCGAAAT 854

```



```

Db      610 CATTGACTATTTTAATGATAATGACCTTAGAAGTATGCTATCTGTAGCTG-AAT 668
Qy      855 TATAGGTACATCTGCTTTTCACTATATGATATTAAGAAAGCTG-AATGACTTAATGTTG 913
Db      669 TATAGGTACATCTGCTTTTCACTATATGATATTAAGAAAGCTGAAATGACTTAATGTTG 728
Qy      914 ATTTTTCCTGTATAGATCTTATCATGTTTTTCATGATTTTAGAAATTAAGTCTTGT 973
Db      729 ATTTTTCCTGTATA-AACTTATCATGTTTTCATGATTTTAGAAATTAAGTCTGCTG- 786
Qy      974 GATATTCGAAAGTGTGAAACTTAAAGTTTATGTTGTACTTTAATCTTGGCATGTTGCT 1033
Db      787 GATATTCGAA--GTGAAACTTAAACTTTATGTTG-ACTTTAATTTCTTGGC-TGTTGCT 842
Qy      1034 CTATGTCCCATTTAA 1049
Db      843 CTATGTCCCATTTAA 858

```

Search completed: January 4, 2005, 13:36:25  
 Job time : 4274.51 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 4, 2005, 11:20:50 ; Search time 16.5 Seconds

(without alignments)

822.216 Million cell updates/sec

Title: US-09-701-618a-2

Sequence: 1 MAGEELNEDYPRVIEHVELA.....LWEPKSNASKVANKGSKS 141

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133.5	18.7	133	2 T41115	sun-cor hormone re
2	110	15.4	184	2 S46808	hypothetical prote
3	83	11.6	207	2 S70534	bhK2.10 protein pr
4	83	11.6	207	2 S70533	bhK2.10 protein pr
5	82	11.5	698	2 S52566	myosin heavy chain
6	81.5	11.4	596	2 T44348	GTP binding protei
7	81.5	11.4	904	2 T03806	hypothetical prote
8	81	11.3	444	2 S64912	probable membrane
9	81	11.3	530	2 G71157	hypothetical prote
10	78.5	11.0	250	2 T22093	hypothetical prote
11	78.5	11.0	508	2 S73430	glycerol kinase gl
12	78.5	11.0	1101	2 T31550	hypothetical prote
13	78.5	11.0	1708	2 AE1866	WD-40 repeat prote
14	78	10.9	630	1 A37097	fimbria - chicken
15	78	10.9	649	2 A31016	hypothetical prote
16	77.5	10.8	2285	2 T12796	probable transglyc
17	77	10.8	466	2 C89820	cysteine1-cRNA syn
18	77	10.8	760	2 T31556	hypothetical prote
19	76.5	10.7	705	2 A71211	probable phosphori
20	76	10.6	446	2 G72731	cell division prot
21	76	10.6	656	2 AH2402	ATP-dependent DNA
22	76	10.6	674	2 A10922	conserved hypotet
23	75.5	10.6	287	2 E82978	tRNA delta-2-isope
24	75.5	10.6	295	2 AB2464	sugar binding prot
25	75.5	10.6	305	2 AF0305	hypothetical prote
26	75.5	10.6	481	2 T14300	hypothetical prote
27	75	10.5	197	2 G72286	transcription regu
28	75	10.5	291	2 E69777	glutamy1-tRNA redu
29	75	10.5	415	1 A48359	

30	75	10.5	424	2 E98008	conserved hypotet
31	75	10.5	992	2 S54396	protein L precursor
32	74.5	10.4	248	2 D71572	hypothetical prote
33	74.5	10.4	312	2 D84956	hypothetical prote
34	74.5	10.4	710	2 S40934	hypothetical prote
35	74.5	10.4	1021	2 T15900	hypothetical prote
36	74.5	10.4	1311	2 A56390	mannosyl-glycoprot
37	74	10.3	204	2 T32902	hypothetical prote
38	74	10.3	465	2 D90426	glucan phosphoryla
39	74	10.3	471	2 S76021	probable trigger t
40	74	10.3	478	2 C59095	hypothetical prote
41	74	10.3	560	2 T32288	plastin, intestina
42	74	10.3	629	2 A56536	hypothetical prote
43	74	10.3	661	2 S67177	hypothetical prote
44	73.5	10.3	441	2 B71816	hypothetical prote
45	73.5	10.3	604	2 T19682	hypothetical prote

## ALIGNMENTS

## RESULT 1

T41115

sun-cor hormone receptor repression complex homolog - fission yeast (Schizosaccharomyces

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C;Accession: T41115

R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.

submitted to the EMBL Data Library, September 1998

A;Reference number: Z21969

A;Accession: T41115

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-133 &lt;LYN&gt;

A;Cross-references: UNIPROT:O74469; EMBL:AL031540; PIDN:CAA20781.1; GSPDB:GN00068; SPDB:.

A;Experimental source: strain 972h-; cosmid c1739

C;Genetics:

A;Gene: SPDB:SPC1739.07

A;Map position: 3

Query Match 18.7%; Score 133.5; DB 2; Length 133;

Best Local Similarity 27.8%; Pred. No. 0.00018;

Matches 37; Conservative 25; Mismatches 54; Indels 17; Gaps 4;

QY 16 EYLSAFB--NSIGAVDEMKTMSV-SRNEILQKDPLEQAVDVSAITLNSMFWYL 71

Db 4 EYSELFEFLNKQDNDVEDVLPKDAESIFELAEGKSELEQAKLYITMSYAINSTLSFY 63

QY 72 ATGQVNPKEHPVQOELEIRIRVVMNRV---KEITDKKKGKLDGGAARVKNALMPEKS 127

Db 64 KLVGIDNSERPVOEQLORVKNYISKIQAEKKNVPTAEVAVN-----TSNAISSSS 114

QY 128 KNASKVANKGSK 140

Db 115 SNRPVAKDAATR 127

## RESULT 2

S46808

hypothetical protein YHR081w - yeast (Saccharomyces cerevisiae)

C;Species: Saccharomyces cerevisiae

C;Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jul-2004

C;Accession: S46808

R;Favell, T.

submitted to the EMBL Data Library, June 1994

A;Description: The sequence of S. cerevisiae cosmid 9205.

A;Reference number: S46795

A;Accession: S46808

A;Molecule type: DNA

A;Residues: 1-184 &lt;PAV&gt;

A;Cross-references: UNIPROT:P38801; EMBL:U10556; NID:G500825; PID:G500829; GSPDB:GN00008;

C;Genetics:

A;Gene: MTPS:YHR081w



```

Query Match 11.4% Score 815; DB 2; Length 596;
Best Local Similarity 23.8% Pred. No. 28;
Matches 29; Conservative 26; Mismatches 32; Indels 35; Gaps 5;

Oy 15 HEYLSAFENSIGAV--DEMLKTMMSVSRNELLOKLDPLEQAKVDLSVAYTLNSMFWYLA 72
Db 437 HDLVESFKSTLEEVVNSLLHVVDSSEKLYEQGDVAE-----WVLS 479
Oy 73 TQGVNPKPE-----HPVQELERIRVYNNRYKEITDKKKAGK-LDRGAASRFVKNAL 122
Db 480 QLVGADPKECILLVNLKIDKLTKEELEDIELKLNKYKLVYISAKGTENLD-----LIINAI 533

Oy 123 WE 124
Db 534 CE 535

RESULT 7
T03806
C:Species: Mycoplasma hyopneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C:Accession: T03806
R:Hu, T.; Minlon, F.C.
Gene 214, 13-23, 1998
A:Title: Molecular analysis of the P97 cilium adhesion operon of Mycoplasma hyopneumoniae
A:Reference number: Z15097; MUID:98391007; PMID:97292120
A:Accession: T03806
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-904 <HSU>
C:Cross-references: UNIPROT:O50187; EMBL:AF012905; NID:g2654768; PIDN:AA032527.1; PID:g27
C:Genetics:
A:Genetic code: SGC3
A:Note: P102
C:Superfamily: Mycoplasma hyopneumoniae hypothetical protein P102

Query Match 11.4% Score 81.5; DB 2; Length 904;
Best Local Similarity 24.4% Pred. No. 46;
Matches 41; Conservative 29; Mismatches 43; Indels 55; Gaps 10;

Oy 9 DYPEVIEHYLSAFENSIGAV-----DEMLKTMMSVSRNELLOKLDPL-EQAK-- 54
Db 418 DLAKDPMKYGQAVALQMGTTIKAVKEFSMPEDQDAKTL---SSNEIKGRVDRFLPAKTVT 474
Oy 55 -VDLVSAVTLNSMFWYLATOG---VNPKEHPVQELERIRVYNNRYKEITDKKKAGKLD 110
Db 475 NLENPSEEVLSK---IYLLNGKYLVDODSEKVGQELKTV-----LEGLSKKNATOKTE 525
Oy 111 RGAASRFVKNALWEPK-----SKNASVYAN-----KGSKS 141
Db 526 -----KNSPTQKKPEVSLAKTTENSAKTVKVSFAFEAKQSQSQS 565

RESULT 8
S64912
C:probable membrane protein YLR080W - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein U2365
C:Species: Saccharomyces cerevisiae
C:Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
C:Accession: S64912
R:Pol, T.M.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64899
A:Accession: S64912
A:Molecule type: DNA
A:Residues: 1-444 <POH>
A:Cross-references: UNIPROT:Q12396; EMBL:Z73252; NID:g1360442; PID:g1360443; GSPDB:GN000
C:Genetics:
A:Gene: MIPS:YLR080W
A:Cross-references: SGD:S0004070

```

A:Map position 12R  
C:Keywords: transmembrane protein  
F:413-429/Domain: transmembrane #status predicted <TM>

Query Match 11.3%; Score 81; DB 2; Length 444;  
Best Local Similarity 27.8%; Pred. No. 22;  
Matches 32; Conservative 20; Mismatches 43; Indels 20; Gaps 5;

OY 13 ETHEVLSAENSISGA--VEBMLKTMMSVSRR--ELLOKLDPLEQAKVDLVSATYLNSFW 68  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 318 ELFEKNKKLEGKIMANDIDPLLRKNNKIYENREBLQRRLP---LDLKTAISDDSFQ 373

OY 69 VVLATGVVSKHPVQGLERIRIVNMVRKEITDKKAGGLDRGAARFVNALW 123  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 374 DFLS---MNAINDRLIKGEQKR-----ODAKLYGQTKGHDEIFSKIYW 416

RESULT 9  
G71157  
hypothetical protein PH0462 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 09-Jul-2004  
A:Accession: G71157  
R:Kawarabayashi, Y.; Sawaeda, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Punahshih, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaean  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: G71157  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-530 <KAM>  
A:Cross-references: UNIPROT:O58213; GB:AP000002; NID:g3236129; PID:BAA29548.1; PID:g3256  
A:Experimental source: strain OT3  
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH0462

Query Match 11.3%; Score 81; DB 2; Length 530;  
Best Local Similarity 25.2%; Pred. No. 27;  
Matches 34; Conservative 29; Mismatches 32; Indels 40; Gaps 8;

OY 10 YPEVHEHYSIAEENGIGAV-----DEMLKTMMSVSR---NELLOKLDPLEQAYD 56  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 401 YP-KIYOETTFVGKSLVALKPSPSGAHNEHLTKEYLSITEKALFNHLSKSIDPSTFSQVR 459

OY 57 LVSATLNSM-----FMVYLATGVNPKENHVKQELERIVVMNVKTEITDKKKGKL 109  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 460 IM---LNSMNNEDEIRIFAMPFLT---SAKPEAKQLLE--KVHVNELKTTID----- 503

OY 110 DRGAASRFVNALWE 124  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 504 ---AFYKFRDIPWE 515

RESULT 10  
T22093  
hypothetical protein F42B11.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
A:Accession: T22093  
R:Lennard, N.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: Z19512  
A:Accession: T22093  
A>Status: preliminary; translated from GB/EMBL/DDBT  
A:Molecule type: DNA  
A:Residues: 1-250 <MW>  
A:Cross-references: UNIPROT:O20334; EMBL:Z66562; PIDD:CAA91466.1; GSPDB:GN00028; CESP:F4,  
A:Experimental source: clone F42B11  
C:Genetics:  
A:Gene: CESP:F42B11.4  
A:Map position: X

A:introns: 20/3; 62/3; 105/1; 153/2; 202/3  
C:Superfamily: tropomyosin I

Query Match  
Best Local Similarity 11.0%; Score 78.5; DB 2; Length 250;  
Matches 28; Conservative 32; Mismatches 46; Indels 13; Gaps 5;

QY 24 SIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDL--VSATYATLSMFVYVATGQVNPKEH 81  
DB 97 NVDSIDHAK--LEAIYNDLFSRLCNLEEEKYDINHITTEETITQINIEVNDLGRKF- 153  
QY 82 PVKPELEIRIVYNNRVEKITDKKXAGLDGASRFYKNAIMPEKSKNA-SKVANKGS 139  
DB 154 -VKSELKKVSKYDKMFKKMAEAKED-----GSKNLRNLTAKVKGESVFTQIANKKS 205

#### RESULT 11

glycerol kinase glpK - Mycoplasma pneumoniae (strain ATCC 29342)  
N:Alternate names: hypothetical protein D09\_ort508  
C:Species: Mycoplasma pneumoniae  
A:Variety: ATCC 29342  
C>Date: 26-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C:Accession: S73430  
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.  
A>Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
A:Reference number: S73430; PMID:97105885; PMID:8948633  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-508 <HIM>  
A:Cross-references: UNIPROT:P75064; EMBL:AE000012; GB:U00089; NID:g1673755; PIDD:AA8575  
C:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
A:Genetics:  
A:Gene: glpK  
A:Genetic code: SGC3  
C:Superfamily: xylulokinase

Query Match  
Best Local Similarity 11.0%; Score 78.5; DB 2; Length 508;  
Matches 32; Conservative 20; Mismatches 54; Indels 19; Gaps 5;

QY 10 YPEIHEVLSAFENSI---GA---VDEMKTMMSSVRNELLQKLDPLEQAKVDLSAYT 62  
DB 298 WOLENQKVYVLESGVFYAGALAKWLRDLSKMTSAESDPYAKLAQKEOEYVFPAPF 357  
QY 63 -LNSMFV-----VYLATGVNPKENHPVKOELERIRVYNN-----RVKEITDKKXGKLD 110  
DB 358 GLGAPYWDASARGAIFGIEANTKREHLVKATLEAIFQANDLTKMAASDLNNSITKIKAD 417  
QY 111 RGAAS 115  
DB 418 GGAACN 422

#### RESULT 12

T31550  
hypothetical protein Y47D3A.26 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T31550  
R:Matthews, L.  
submitted to the EMBL Data Library, October 1999  
A:Reference number: Z21043  
A:Accession: T31550  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1101 <MII>  
A:Cross-references: UNIPROT:Q9U2C1; EMBL:AL117202; PIDD:CA57698.1; CESP:Y47D3A.26  
C:Experimental source: clone Y47D3A  
A:Genetics:  
A:Gene: CESP:Y47D3A.26

A:introns: 135/2; 169/3; 252/3; 287/2; 320/3; 387/2; 441/3; 637/3; 681/3; 786/3; 842/3;  
C:Superfamily: hypothetical protein YJL074C

Query Match  
Best Local Similarity 11.0%; Score 78.5; DB 2; Length 1101;  
Matches 29; Conservative 14; Mismatches 35; Indels 31; Gaps 3;

QY 14 IHEVLSAFENSIQAVDEMLK-----TMMSSVRNELLQKLDPLEQAKVDLSAYTL 63  
DB 13 IDERLQTELENKEDLKEYQKLDKTRKSEVETMDNTKKAIRKKTLDQKQVFL----- 66  
QY 64 NSMFVYVATGQVNPKEHHPVKOELERIRVYNNRVEKITDKKXGKLDG 112  
DB 67 -----NQKDNNTKSQLNDVIAMETAKL--TDKKKLESIGRG 100

#### RESULT 13

AE1866  
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. strain PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AE1866  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; PMID:21595285; PMID:11759840  
A:Accession: AE1866  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1708 <KUN>  
A:Cross-references: UNIPROT:Q8VZ12; GB:BA000019; PIDD:BA87436.1; PIDD:g17129823; GSPDB:C  
A:Experimental source: strain PCC 7120  
A:Genetics:  
A:Gene: al10478

Query Match  
Best Local Similarity 11.0%; Score 78.5; DB 2; Length 1708;  
Matches 36; Conservative 31; Mismatches 51; Indels 35; Gaps 5;

QY 5 EINEDYPEIHEVLSAFENSIQAVDEMLKTMMSVSRNELLQKLDPLEQAK----- 54  
DB 899 QLIHLYVLDLIRYLOQSSSLQAOGLDQRRKVOQSGEIRAKSELQKQSKLTDTHPQ 958  
QY 55 --VDLSAYTLNSMFVYVATGQVNPKEHHPVKOELERIRVYNNRVEKIT-----DKKXGK 108  
DB 959 QGLDLYT-----ELRELKREBELTQLEIBQRLGEL--KEKELTYQLAESQKQK 1005  
QY 109 LDRGASRFYKNAIMPEKSKNAASVYANKGSKS 141  
DB 1006 LSEAKLNRSIKIAL-----TASCLAILGLSVS 1032

#### RESULT 14

A37097  
fimbria - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 01-Feb-1991 #sequence\_revision 14-Jul-1994 #text\_change 09-Jul-2004  
C:Accession: A37097; S12080  
R:de Arruda, M.V.; Watson, S.; Lin, C.S.; Leavitt, J.; Matsuda, P.  
U. Cell Biol. 111, 1069-1079, 1990  
A>Title: Fimbria is a homologue of the cytoplasmic phosphoprotein plastin and has domain  
A:Reference number: A37097; PMID:90361735; PMID:2391360  
A:Accession: A37097  
A:Molecule type: mRNA  
A:Residues: 1-630 <DEA>  
A:Cross-references: UNIPROT:P19179; GB:X52562; NID:962889; PIDD:CA36796.1; PIDD:962890  
C:Superfamily: plastin; alpha-actinin actin-binding domain homology; calmodulin repeat h  
C:Keywords: actin binding; blocked amino end; calcium binding; EF hand  
F/11-43/Domain: calmodulin repeat homology <EF1>  
F/51-83/Domain: calmodulin repeat homology <EF2>  
F/121-372/Domain: alpha-actinin actin-binding domain homology <ACT1>

F;395-621/Domain: alpha-actinin actin-binding domain homology <ACT2>

Query Match	Score	DB	Length
10.9%	78	1	630

Best Local Similarity 25.7%; Pred. No. 60;

Matches 29; Conservative 21; Mismatches 45; Indels 18; Gaps 4;

31 MLKTMMSVRNELLQKDLPLBQAKVD--LVSA YTLNSMFWVYLATQGVNPKHPVKQEL 87

```
Db      1 MENVTTISREELLELRFAFNKIDIDNSGYVSDYELQDLF-----KEASLPLPG 49
```

88 ERIRVYNNRVEITDDKKAGKLDRGASRFVYNALWEPKSKNASKVANKGSK 140

Db 50 YKREIEKI PAVTDSNKDGKIN--FEEFV-SLIQELSKOVSKSYRSINK 98

**RESULT 15**

**S41016**

hypothetical protein T07C4.3 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

CjDate: 06-Jan-1995 #sequence 06-Jan-1995 #text\_change 09-Jul-2004  
CjAccession: 641016

C;Accession: S41016

Submitted to the EMML Data Library, January 1994

submitted to the EMBL Data Library, January 1994  
A: Reference number: S41014

A;Reference number: S41014  
A:Acceassion: S41016

A;Accession: S41016  
A;Status: preliminary

A:Status: preliminary  
A:Molecule type: DNA

A:molecule type: DNA  
A:Residue: 1-649 <BER>

A:Cross-references: UN1

A/Genetics: UN1FR01:Q222250; EMBL:Z23443; MID:G108/051; FID:G44303

A; Introns:

$\frac{c}{\lambda_{00}} = \frac{c}{\lambda_0} + \frac{\beta^2 c}{\lambda_0(1-\beta)} = \frac{c}{\lambda_0} + \frac{v^2}{\lambda_0(1-\beta)}$

Query Match	10.9%;	Score 78;	DB 2;	Length 649;
Best Local Similarity	28.4%;	Pred. No. 62;		

Best Local Similiarily 28.4%, Fred: NO. 82,  
Matches 40: Conservative 20: Mismatches

malicious 40; combi valve 20; misbehavior 80; insects 10; bags 10

11 PVEIHEYL-SAFE---NSIGAVDEMLKTIMSVSRNELLOKLDPLBQAKVDLSAYTILNSM 66

D6 281 PIVLTESIDAFEAIVIQSVSHADATLSMTQNASSEAKVHATPPLPAERKTMVSAELANGK 340

QY 67 FVWYLATQ---GVNPKKEHPVK---QELERIRVYNNRVKEITDKKKAGKLDRGAAASRFVK 119

Db 341 PIKHLIEKFDAGVNF AEHKPKSIYAQVLEEIGGSAPRVDEVF---SASKKEEHAETE-VT 396

QY 120 NALWEPKSKNASKVANKGSK 140

Db 397 NVIFRSTSTHSSIIAN-GKEQ 416

Search completed: January 4, 2005, 13:45:13  
Job time : 23.5 sec8

Job time : 23.5 sec

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 4, 2005, 08:33:30 ; Search time 63.5 Seconds

(without alignments)  
1277.604 Million cell updates/sec

Title: US-09-701-618A-2  
Perfect score: 715  
Sequence: 1 MAGEEINEDYPTVEIHEYL...LMEPKSKNASKVANKSKSKS 141

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues  
Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Uniprot 02: \*  
1: uniprot\_trembl: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	141	2	013901 homo sapien
2	654	91.5	141	2	035473 m small uni
3	652	91.2	141	2	071500 Q71500 cricetulus
4	646	90.3	141	2	061368 mus musculu
5	549	76.8	111	2	071861 homo sapien
6	186.5	26.1	141	2	086F82 schistosoma
7	161	22.5	204	2	06B422 debaryomyce
8	150.5	21.0	217	2	093VA9 arabidopsis
9	139.5	19.5	169	2	070127 anopheles g
10	133.5	18.7	133	2	074469 schizosacch
11	133	18.6	194	2	06C2H9 yarrowia 11
12	132	18.5	207	2	06YX09 cryza sativ
13	132	18.5	207	2	BAC99871 bac99871 cryza sat
14	128.5	18.0	133	2	09N3J4 caenorhabdi
15	126.5	17.7	159	2	09VXL4 drosophila
16	121	16.9	450	2	07SER2 neurospora
17	119	16.6	149	2	081L18 plasmodium
18	110	15.4	184	1	YH01 YEAST
19	109	15.2	149	2	06FUI1 saccharomyc
20	106.5	14.9	177	2	06CMK3 kluyveromyc
21	94.5	13.2	654	2	081J34 arabidopsis
22	90	12.6	269	2	08TQ04 pyrococcus
23	89	12.4	398	2	08IKD8 plasmodium
24	87.5	12.2	915	2	032863 mycoplasma
25	86	12.0	1454	2	06CT15 kluyveromyc
26	85	12.0	244	2	09HGN2 schizosacch
27	85	11.9	150	2	0830G2 enterococcu
28	85	11.9	336	2	07RXH8 neurospora
29	85	11.9	1208	2	09VXU1 drosophila
30	85	11.9	1398	2	09VXU2 drosophila
31	84.5	11.8	530	2	06MDN2 parachlamy

32	84.5	11.8	530	2	CAP23317	Cap23317 parachlam
33	84	11.7	2867	1	RBP2 PLAVB	000799 plasmodium
34	83.5	11.7	129	2	0977N0	0977N0 unclutured
35	83.5	11.7	2511	2	081L44	081L44 plasmodium
36	83	11.6	207	2	044732	044732 borella bu
37	83	11.6	207	2	044733	044733 borella bu
38	83	11.6	1249	2	08DT19	08dt19 streptococc
39	83	11.6	2104	2	08XW24	08xw24 clostridium
40	82.5	11.5	642	2	09S70	09s70 mus musculu
41	82.5	11.5	642	2	09D0F1	09d0f1 mus musculu
42	82.5	11.5	1104	2	06BH46	06bh46 debaryomyce
43	82	11.5	395	2	06CQF7	06cqt7 kluyveromyc
44	82	11.5	455	2	09SG79	09sg79 arabidopsis
45	82	11.5	698	2	Q91208	Q91208 oncorhynch

## ALIGNMENTS

RESULT 1	ID	PRELIMINARY	PRT	141 AA.
013901	Q13901			
AC	Q13901			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-OCT-2004 (TrEMBLrel. 28, Last annotation update)			
DE	CID protein (Nuclear DNA-binding protein).			
GN	Name=CID;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
NC	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Bone marrow; Lung, and Uterus;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Krausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stampleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshitoki S., Carninci P., Prange C.,			
RA	Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,			
RA	Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,			
RA	Jones S.J., Watra M.A.,			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Bone marrow;			
RA	Strausberg R.,			
RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Uterus;			

RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=C57BL/6J;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X95592; CA64845.1; -  
 DR EMBL; BC009584; AA09589.1; -  
 DR EMBL; BC009589; AA09589.1; -  
 DR EMBL; BC016284; AA016284.1; -  
 DR EMBL; BC005235; AA05235.1; -  
 DR GO; GO:0005634; C:nucleus; TAS.  
 DR GO; GO:0003677; F:DNA binding; TAS.  
 DR InterPro; IPR011082; CID.  
 DR Pfam; PF07493; CID; 1.  
 DR DNA-binding.  
 KW KW  
 SQ SEQUENCE 141 AA; 16019 MW; 9976A3BBD562D063 CRC64;  
 Query Match 100.0%; Score 715; DB 2; Length 141;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-56;  
 Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAGEINEDYVPEIHEHLSAFENSIGAVDEMLKTMMSSVRNELLQKLDPEQAKYDLVSA 60  
 DB 1 MAGEINEDYVPEIHEHLSAFENSIGAVDEMLKTMMSSVRNELLQKLDPEQAKYDLVSA 60  
 QY 61 YTTNSFWVYLATQGVNPKHPKQBLERIRVYNNVKEITDKKAGKLDGAAARPYKN 120  
 DB 61 YTTNSFWVYLATQGVNPKHPKQBLERIRVYNNVKEITDKKAGKLDGAAARPYKN 120  
 QY 121 ALMEPSKNSKVKVANGSKS 141  
 DB 121 ALMEPSKNSKVKVANGSKS 141  
 RESULT 2  
 ID 035473 PRELIMINARY; PRT; 141 AA.  
 AC 035473;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last annotation update)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Small unique nuclear receptor co-repressor (Nuclear DNA-binding  
 protein) (Mus musculus 10 days neonate skin cDNA, RIKEN full-length  
 enriched library, clone:4732439J08 product:nuclear DNA binding  
 protein, full insert sequence) (Mus musculus 12 days embryo embryonic  
 body between diaphragm region and neck cDNA, RIKEN full-length  
 enriched library, clone:9430096A21 product:nuclear DNA binding  
 protein, full insert sequence).  
 DE Name=Cid; Synonyms=SUN-COR;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98070763; PubMed=9405624;  
 RA Zamir I., Dawson J., Lavinsky R.M., Glass C.K., Rosenfeld M.G.,  
 Lazar M.A.;  
 RT "Cloning and characterization of a corepressor and potential component  
 of the nuclear hormone receptor repression complex.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:14400-14405(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6J;  
 RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;  
 RX MEDLINE=23386257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Gronow L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner L., Shmelen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bat N.K.,  
 Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 Brownstein W.J., Usdin T.B., Toshiyuki S., Carninci P., Pringle C.J.,  
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhally S.J.,  
 Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huij S.W.,  
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Rodriguez R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 Krzywinski M.I., Skaleka U., Smallos D.E., Schnerch A., Schein J.E.,  
 Jones S.J., Maira M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6J;  
 RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency long-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RA The FANTOM Consortium;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=20499374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subcloning of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,  
 Yamamoto R., Matsumoto H., Nakamura S., Hazama M., Nishine T., Harada A.,  
 Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe K.,  
 Ozaki Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multiplexed capillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [9]  
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carminci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,  
 RA Hori F., Imocani K., Iehi Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai K., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ono M., Ono N., Okazaki Y.,  
 RA Saito R., Satoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF031426; AAC53520.1; -;  
 DR EMBL; BC005436; AAH05436.1; -;  
 DR EMBL; AK028702; BAC26075.1; -;  
 DR EMBL; AK035169; BAC28967.1; -;  
 DR MGI; MGI:1927354; ClD.  
 DR GO; GO:0017053; C:transcriptional repressor complex; IDA.  
 DR GO; GO:0003677; F:DNA binding; IDA.  
 DR GO; GO:0016922; F:ligand-dependent nuclear receptor interacto. . .; IDA.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR GO; GO:0003714; F:transcription corepressor activity; IDA.  
 DR GO; GO:0016481; F:negative regulation of transcription; IDA.  
 DR InterPro; IPR011082; ClD.  
 DR Pfam; PF07493; ClD; 1.  
 DR DNA-binding; Receptor.  
 KW NCBI\_TaxId=10090;  
 SQ SEQUENCE 141 AA; 15945 MW; C9AF68033A0FD04E CRC64;  
 Query Match 91.5%; Score 654; DB 2; Length 141;  
 Best Local Similarity 90.7%; Pred. No. 2.3e-50;  
 Matches 127; Conservative 8; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 MAGEINEDYVPEIHETLSAFENSIGAVDEMKTMSVSNELLOKLDPLEQAKVDLVA 60  
 DB 1 MAGEBNMEDYVPEIHESLTALSSLGAVDMLKTMAVSNELLOKLDPLEQAKVDLVA 60  
 QY 61 YTLNSMFVWYLAQGVNPKHPVKQELERIRVYNNRVEKITDDKKKAGLDGGAASRFVK 120  
 DB 61 YTLNSMFVWYLAQGVNPKHPVKQELERIRVYNNRVEKITDDKKKAGLDGGAASRFVK 120  
 QY 121 ALMEPKSKNASKVANKGSK 140  
 DB 121 ALMEPKSKNASKVANKGSK 140  
 RESULT 3  
 Q7TSUO PRELIMINARY; PRT; 141 AA.  
 AC Q7TSUO;  
 DT 01-OCT-2003 (TREMBLrel. 25, Created)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Nuclear DNA-binding protein ClD  
 OS Cricetus griseus (Chinese hamster)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetulus.  
 OX NCBI\_TaxId=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Guang L., Masabumi S., Maru Y.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY302220; AAP43113.1; -;  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR011082; ClD.  
 DR Pfam; PF07493; ClD; 1.  
 DR DNA-binding.  
 KW NCBI\_TaxId=10029;  
 SQ SEQUENCE 141 AA; 15945 MW; 0CFB6B47764B6566 CRC64;  
 Query Match 91.2%; Score 652; DB 2; Length 141;  
 Best Local Similarity 91.4%; Pred. No. 3.4e-50;

Matches 126; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 MAGEINEDYVPEIHETLSAFENSIGAVDEMKTMSVSNELLOKLDPLEQAKVDLVA 60  
 DB 1 MAGEBNMEDYVPEIHESLTALSSLGAVDMLKTMAVSNELLOKLDPLEQAKVDLVA 60  
 QY 61 YTLNSMFVWYLAQGVNPKHPVKQELERIRVYNNRVEKITDDKKKAGLDGGAASRFVK 120  
 DB 61 YTLNSMFVWYLAQGVNPKHPVKQELERIRVYNNRVEKITDDKKKAGLDGGAASRFVK 120  
 QY 121 ALMEPKSKNASKVANKGSK 140  
 DB 121 ALMEPKSKNASKVANKGSK 140  
 RESULT 4  
 Q61368 PRELIMINARY; PRT; 141 AA.  
 AC Q61368;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE ClD protein.  
 GN Name=Cld; Synonyms=Cld;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NMRI; TISSUE=Ascites tumor;  
 RX MEDLINE=98136194; PubMed=9469821;  
 RA Nehls P., Kecke T., Guderath R., Spiess E., Glaeser T., Rothbarth K.,  
 RA Stammer H., Werner D.;  
 RT "CDNA cloning, recombinant expression and characterization of  
 RT polypeptides with exceptional DNA affinity."  
 RL Nucleic Acids Res. 26:1160-1166(1998).  
 DR EMBL; X95591; CAA64844.1; -;  
 DR MGI; MGI:1927354; ClD.  
 DR GO; GO:0017053; C:transcriptional repressor complex; IDA.  
 DR GO; GO:0016922; F:ligand-dependent nuclear receptor interacto. . .; IDA.  
 DR GO; GO:0003677; F:DNA binding; IDA.  
 DR GO; GO:0016922; F:ligand-dependent nuclear receptor interacto. . .; IDA.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR GO; GO:0003714; F:transcription corepressor activity; IDA.  
 DR GO; GO:0016481; F:negative regulation of transcription; IDA.  
 DR InterPro; IPR011082; ClD.  
 DR Pfam; PF07493; ClD; 1.  
 KW NCBI\_TaxId=10090;  
 SQ SEQUENCE 141 AA; 16045 MW; C9AF68033A097161 CRC64;  
 Query Match 90.3%; Score 646; DB 2; Length 141;  
 Best Local Similarity 90.0%; Pred. No. 1.2e-49;  
 Matches 126; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 MAGEINEDYVPEIHETLSAFENSIGAVDEMKTMSVSNELLOKLDPLEQAKVDLVA 60  
 DB 1 MAGEBNMEDYVPEIHESLTALSSLGAVDMLKTMAVSNELLOKLDPLEQAKVDLVA 60  
 QY 61 YTLNSMFVWYLAQGVNPKHPVKQELERIRVYNNRVEKITDDKKKAGLDGGAASRFVK 120  
 DB 61 YTLNSMFVWYLAQGVNPKHPVKQELERIRVYNNRVEKITDDKKKAGLDGGAASRFVK 120  
 QY 121 ALMEPKSKNASKVANKGSK 140  
 DB 121 ALMEPKSKNASKVANKGSK 140  
 RESULT 5  
 Q7L8E1 PRELIMINARY; PRT; 111 AA.  
 AC Q7L8E1;  
 DT 05-JUL-2004 (TREMBLrel. 27, Created)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)

DE BA19501.1 (Novel protein similar to CID DNA-binding protein).  
 GN Name=BA19501.1;  
 OS Homo sapiens (human).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RX NCB1\_Taxid=9606;  
 RP SEQUENCE FROM N.A.  
 RA Lawlor S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL354750; CADI3204.1;  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR011082; CID.  
 DR Pfam; PF07493; CID; 1.  
 KW DNA-binding.  
 SQ SEQUENCE 111 AA; 12646 MW; F563BA631CECEE1 CRC64;

Query Match 76.8%; Score 549; DB 2; Length 111;  
 Best Local Similarity 98.2%; Pred. No. 3.9e-41;  
 Matches 109; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 31 M L K T M S V S R N E L L Q L D L E Q A K V D L V S A Y T L N S M F W Y L A T G V N P K E H P V K O E L E R I 90  
 Db 1 M L K T M S V S R N E L L Q L D L E Q A K V D L V S A Y T L N S M F W Y L A T G V N P K E H P V K O E L E R I 90  
 Qy 91 R V Y M N R V E I T D K K A G L D R G A S R F Y K N A L M E P K S K N S K V A N K G S K S 141  
 Db 61 R V Y M N R V E I T D K K A G L D R G A S R F Y K N A L M E P K S K N S K V A N K G S K S 111

## RESULT 6

086F82 PRELIMINARY; PRT; 141 AA.

DT 01-JUN-2003 (Tremblrel. 24, Created)  
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
 DE 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
 OS Schistosoma japonicum (Blood fluke).  
 OC Schistosoma japonicum (Blood fluke).  
 CC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;  
 CC Schistosomatidae; Schistosomatidae; Schistosoma.  
 RX NCB1\_Taxid=6182;  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=22879925; PubMed=12973349;  
 RA Hu W., Yan Q., Shen D.K., Liu F., Zhu Z.D., Song H.D., Xu X.R.,  
 RA Wang Z.J., Rong Y.P., Zeng L.C., Wu J., Zhang X., Wang J.J., Xu X.N.,  
 RA Xue C.L., Feng Z., Chen Z., Han Z.G., Brindley P.O., McKann D.P.,  
 RT "Evolutionary and biomedical implications of a Schistosoma japonicum  
 complementary DNA resource."  
 RL Nat. Genet. 35:139-147(2003).  
 DR EMBL; AY22982; AAP06005.1;  
 DR InterPro; IPR011082; CID.  
 DR Pfam; PF07493; CID; 1.  
 SQ SEQUENCE 141 AA; 15966 MW; DBEFB20FA1894BA0 CRC64;

Query Match 26.1%; Score 186.5; DB 2; Length 141;  
 Best Local Similarity 29.1%; Pred. No. 9.6e-09;  
 Matches 39; Conservative 33; Mismatches 59; Indels 3; Gaps 2;

Qy 1 M A G E I N E D Y P V E H E T L S A F E N S I G A V D E M L K T M S V S R N E L L Q L D L E Q A K V D L V S A Y T L N S M F W Y L A T G V N P K E H P V K O E L E R I 60  
 Db 1 M G S S I D E L P K E I S S O L V S F S E A T D V E Q L V N K I S S F S N N S - S N E V S G L D V I S E L S L C 59  
 Qy 61 Y T L N S M F W Y L A T G V N P K E H P V K O E L E R I Y M N R V E I T D K K A G L D R G A S R F Y K N A L M E P K S K N S K V A N K G S K S 141  
 Db 60 Y A N N A F F W L R C N G V E T O S H P I M G E L D V N M A L K R C S L V E R V S A R L T L D K E A T R F V 118  
 Qy 119 K N A L M E P K S K N S K 132  
 Db 120 K N A L M S A H T T K K 133

## RESULT 7

06B372 PRELIMINARY; PRT; 204 AA.

ID 06B372  
 AC 06B372;  
 DT 01-OCT-2004 (Tremblrel. 28, Created)  
 DT 01-OCT-2004 (Tremblrel. 28, Last sequence update)  
 DE 01-OCT-2004 (Tremblrel. 28, Last annotation update)  
 GN ORFNames=DEHA0F27797g;  
 OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.  
 RX NCB1\_Taxid=4959;  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CBS767;  
 RG GENOLEYURES;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,  
 RA Gottard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
 RA Battey S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,  
 RA Boissiere A., Boyer J., Catolico L., Comandolieri F., de Daruvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantreya F., Hennequin C., Jaubaux N., Joyet P., Kachouri R.,  
 RA Kerrist A., Kozul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicland J.M., Nikolski M., Oztas S., Ozler-Kalogeropoulos O.,  
 RA Pellenz S., Pocher S., Richard G.F., Straub M.L., Suleau A.,  
 RA Swemene D., Tekala F., Wesolowski-Louvel M., Weithof E., Wirth B.,  
 RA Zeniou-Weyer M., Zivanovic I., Bolotin-Pukhata M., Thierry A.,  
 RA Bouchier C., Caudon B., Scarpelli C., Gaillardin C., Weissenbach J.,  
 RA Wincker P., Souciet J.L., Nature 430:35-44(2004).  
 RT "Genome evolution in yeasts."  
 RL Nature 430:35-44(2004).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS767;  
 RA Genoscope;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; CR382138; CAG89900.1;  
 SQ SEQUENCE 204 AA; 22569 MW; 724A5F9FC34D0C9C CRC64;

Query Match 22.5%; Score 161; DB 2; Length 204;  
 Best Local Similarity 32.7%; Pred. No. 2.7e-06;  
 Matches 37; Conservative 27; Mismatches 45; Indels 4; Gaps 2;

Qy 14 I H E Y L S A F E N S I G A V D E M L K T M S V S R N E L L Q L D L E Q A K V D L V S A Y T L N S M F W Y L A T G V N P K E H P V K O E L E R I 72  
 Db 7 V N L F V Q S L D N S V D Q L E K L E P L K S L E K T A A S D S O V E R I K I Y N N S Y V L I S I L S Y L K 66  
 Qy 73 T O G V N P K E H P V K O E L E R I Y M N R V E I T D K K A G L D R G A S R F Y K N A L M E P K S K N S K V A N K G S K S 141  
 Db 67 T L G I N T D Q H P I M K E L T R I K I Y M K R Y E L E A K L A S K O T S K E D A R A A R F I O N T L 119

## RESULT 8

093VA9 PRELIMINARY; PRT; 217 AA.

ID 093VA9  
 AC 093VA9;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DE 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
 DE AT5g25080/711H3.90 (Hypothetical protein).  
 OC Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 RX NCB1\_Taxid=3702;  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Dale J.M., Goldemich A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,

RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Saitou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22088475; PubMed=12093376;  
 RA Haas B.U., Volfovsky N., Town C.D., Troughan M., Alexandrov N.,  
 RA Feldman K.A., Flavell R.B., White O., Salzberg S.L.;  
 RT "Full-length messenger RNA sequences greatly improve genome  
 RT annotation.";  
 RL Genome Biol. 3:RESEARCH0029-RESEARCH0029 (2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Brover V., Troughan M., Alexandrov N., Lu Y.-P., Flavell R.,  
 RA Feldmann K.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,  
 RA Banh J., Bowser L., Carninci P., Chung M.K., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlman-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T.,  
 RA Saitou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yu G., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY050469; AAK91482.1; -  
 DR EMBL: AY087408; AAM64957.1; -  
 DR EMBL: AF378870; AAK55673.1; -  
 DR InterPro: IPR011082; CID: 1.  
 DR Pfam: PF07493; CID: 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 217 AA; 23748 MW; DB4F167BD773DB4 CRC64;  
 SO  
 Query Match 21.0%; Score 150.5; DB 2; Length 217;  
 Best Local Similarity 27.6%; Pred. No. 2.5e-05;  
 Matches 37; Conservative 35; Mismatches 47; Indels 15; Gaps 2;  
 QY 22 ENSGAVDE-----MLKTMVSRRNELLQKLDPLBOAKVDLSAYTNSMFWYL 71  
 DB 18 ESNLEANNQTLAYLKEKLPQLEQMLTAEPVLAAMPPLQAKTMHLAETTLTTLRL 77  
 QY 72 ATGCVNKEHPVQELERIRYNNRVKEITDKK-----AGKLDGGAARFVKNALWEPK 126  
 DB 78 RCTGVDPDDHRRVSEIRINIVYREKFOKCVQDSKGPLRPTVLNRQATRPRIHSLPDLT 137  
 QY 127 SKKASKANKGSKK 140  
 DB 138 STOKQSIKDSKGE 151  
 RESULT 9  
 QY0127 PRELIMINARY; PRT; 169 AA.  
 AC QY0127;  
 DT 01-MAR-2004 (TREMBLrel. 26, Created)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE AGCP3387 (Fragment).  
 GN Name=agCG52216; ORFNames=ENSGA00000016976;  
 OS Anopheles gambiae str. PEST.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 OX NCBI\_TaxID=180454;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 -1- CAUTION: The sequence shown here is derived from an

CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL, AAB0100807; EAA04511.1; -  
 FT NON\_TER 1 1  
 SQ SEQUENCE 169 AA; 19590 MW; 69C2764F057A93C2 CRC64;  
 Query Match 19.5%; Score 139.5; DB 2; Length 169;  
 Best Local Similarity 30.7%; Pred. No. 0.00018;  
 Matches 35; Conservative 28; Mismatches 48; Indels 3; Gaps 3;  
 QY 22 ENSGAVDEMLKTMVSRRN-ELLQKLDPLBOAKVDLSAYTNSMFWYLATG-VNPK 79  
 DB 37 ENLMQAIQOVTLANVRREYKYPNTAEKVRFTLCAFCVNSLFWMEKMLGRNPTV 96  
 QY 80 EHPVQELERIRYNNRVKEITDK-KKAGLDRGAARFVKNALWEPKSKKSK 132  
 DB 97 MDDIKADLRVREMKRLQTHDNLTRPRLDQPAAGRFVAGLDYANNAESDK 150  
 RESULT 10  
 QY04469 PRELIMINARY; PRT; 133 AA.  
 AC QY04469;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE SPC139.07 protein.  
 GN Name=SPC139.07;  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972h-;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sguros J., Peat N., Hayles J., Baker S., Baitham D., Bowman S.,  
 RA Brooke K., Brown D., Brown S., Chillingworth T., Churcher C.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Gobie A., Hamlin N., Harris D., Hildalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle B.J., Hunt S., Jagsels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S., Warren T., Whitehead S.,  
 RA Woodward J., Voickaeert G., Aert R., Robben J., Grymoprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schafer M., Muller-Auer S.,  
 RA Gabel C., Fuchs M., Dusterhoft A., Fritzc C., Holzer E., Moestl D.,  
 RA Hilbert H., Borzym K., Langer I., Beck A., Leirach H., Reinhardt R.,  
 RA Pohl T.M., Eger P., Zimmermann W., Wedler H., Wandt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelare V., Mortier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 DR EMBL: AL031540; CAA20781.1; -  
 DR PIR: T4115; T4115.  
 DR GeneDB Spombe; SPC139.07; -  
 DR InterPro: IPR011082; CID: 1.  
 DR Pfam: PF07493; CID: 1.  
 SQ SEQUENCE 133 AA; 15068 MW; DD333501508D60F1 CRC64;  
 Query Match 18.7%; Score 133.5; DB 2; Length 133;  
 Best Local Similarity 27.8%; Pred. No. 0.00047;  
 Matches 37; Conservative 25; Mismatches 54; Indels 17; Gaps 4;

Qy 16 EYLSAFE---NSIGAVDEMLKTMSV--SRNELLOKLDPLFOAKYDVLVSAYTLNSMFWVYL 71  
D 4 EYSLFPERLNKQDNLNDVDLKPDLKDAESITFELAGSELEOALYITMSFAINSTLYSFY 63  
Qy 72 ATQGVNPKFHVKQELERIRVYNNRV---KEITDKKAGKLDGRGAASRFVKNALMEPKS 127  
D 64 KUNGIDASERPVMOELQVRKVYISKIQOAKKNVPTEAVN-----TSNAALSSSS 114  
Qy 128 KNASKVANKGSKS 140  
D 115 SNRPKVAKDAATR 127

RESULT 11  
Q6C2H9  
ID 06C2H9 PRELIMINARY; PRT; 194 AA.  
AC 06C2H9;  
DT 01-OCT-2004 (TrEMBLrel. 28, Created)  
DT 01-OCT-2004 (TrEMBLrel. 28, last sequence update)  
DE Similar to DEHA0F27797g Debaryomyces hansenii.  
GN ORFNames=YALIOF07755g;  
OS Yarrowia lipolytica (Candida lipolytica).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Dipodascaceae; Yarrowia.  
OX NCBI\_TaxID=4952;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIB99;  
RG GENOLEVURES;  
RA Lafont B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
RA Goffard N., Frangoul L., Aigle M., Anthonard V., Babour A., Barbe V.,  
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykaeten C.,  
RA Boistame A., Boyer J., Catcolico L., Confanier F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
RA Hantraye F., Henequin C., Jauniaux N., Joyet P., Kachouri R.,  
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Nicoud J.M., Nikolski M., Oztas S., Ozler-Kalogeropoulos O.,  
RA Pellenz S., Peltier S., Richard G.F., Straub M.L., Suleau A.,  
RA Swemene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
RA Zentou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,  
RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,  
RA Winkler P., Souciet J.L.,  
RT "Genome evolution in yeasts."  
RL Nature 430:35-44(2004).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIB99;  
RA Genoscope;  
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; CR382132; CAG7940.1; -.  
SQ SEQUENCE 194 AA; 21520 MW; D67D7DF8068EB3B CRC64;

Query Match 18.6%; Score 133; DB 2; Length 194;  
Best Local Similarity 24.8%; Pred. No. 0.0008;  
Matches 38; Conservative 31; Mismatches 58; Indels 26; Gaps 4;  
Qy 13 EIHVYLSAFNSIGAVDEMLKTMSVSRNELLOKLDPLFOAKYDVLVSAYTLNSMFWVYL 72  
D 4 QIEVLELSHNLQDVTSLEQIKMIDFKGVAQ-LPPLLEQAFYSKLAVYTNAMFAFL 62  
Qy 73 TQGVNPKFHVKQELERIRVYNNRVKEITDK-----KKAGLDGGAASRFV----- 118  
D 63 ASGGDPTHTHIMDLDRVKTMYGVAHAEKGPAPAKDEKTKVDVPAKRITIFACTERA 122  
Qy 119 --KNALMEPKS-----KNASKVANKGSKS 140  
D 123 ISADSIKEPPTSDAAEAFKLDVTKSASKDKK 155

RESULT 12

Q6XY09  
ID 06XY09 PRELIMINARY; PRT; 207 AA.  
AC 06XY09;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, last sequence update)  
DE Hypothetical protein OSJNB0091D16.5.  
GN Name=OSJNB0091D16.5;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Eukaryota; Viridiplantae; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzae; Oryza.  
OX NCBI\_TaxID=39947;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Sasaki T., Matsumoto T., Katayose Y.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP005731; BAC99871.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 207 AA; 22203 MW; 0B0161AD8190E071 CRC64;

Query Match 18.5%; Score 132; DB 2; Length 207;  
Best Local Similarity 27.9%; Pred. No. 0.0011;  
Matches 38; Conservative 33; Mismatches 53; Indels 12; Gaps 4;  
Qy 18 LSAFENSIGA---VDEMLKTMSVSRN--ELLQKLDPLFOAKYDVLVSAYTLNSMFWVYL 72  
D 18 VSAEDTTLAAESVGDHLEMAVAAGDPDAIAELPPLRARAPLAKMAKAASLFAVRLR 77  
Qy 73 TQGVNPKFHVKQELERIRVYNNRVKEITDKKA-----GKLDGGAASRFVKNAL--WEP 125  
D 78 CSGVDDEHPKKEPRLTMEKLNRPEDMKAPLPTTTVTOAAAFIGHSLPHLTT 137  
Qy 126 KSNASKVANKGSKS 141  
D 138 DQKRSWQAISSRGGGS 153

RESULT 13  
BAC99871  
ID BAC99871 PRELIMINARY; PRT; 207 AA.  
AC BAC99871;  
DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, last sequence update)  
DE Hypothetical protein OSJNB0091D16.5.  
GN OSJNB0091D16.5.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzae; Oryza; Oryza sativa.  
OX NCBI\_TaxID=39947;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Katayose Y.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC  
clone:OSJNB0091D16.";  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP005731; BAC99871.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 207 AA; 22203 MW; 0B0161AD8190E071 CRC64;

Query Match 18.5%; Score 132; DB 2; Length 207;  
Best Local Similarity 27.9%; Pred. No. 0.0011;  
Matches 38; Conservative 33; Mismatches 53; Indels 12; Gaps 4;  
Qy 18 LSAFENSIGA---VDEMLKTMSVSRN--ELLQKLDPLFOAKYDVLVSAYTLNSMFWVYL 72  
D 18 VSAEDTTLAAESVGDHLEMAVAAGDPDAIAELPPLRARAPLAKMAKAASLFAVRLR 77  
Qy 73 TQGVNPKFHVKQELERIRVYNNRVKEITDKKA-----GKLDGGAASRFVKNAL--WEP 125

DB 78 CSGVDPEHPIKKEPERLSLWEEKLNPFEDWDKAPLPTTNTVTAARFISLPHLTT 137  
 QY 126 KSRKASVANKGSKS 141  
 DB 138 DOKRSMQAIRSGEGS 153

RESULT 14  
 Q9N3J4 PRELIMINARY; PRT; 133 AA.  
 ID Q9N3J4; 01-OCT-2000 (TReMBLrel. 15, Created)  
 AC 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TReMBLrel. 25, Last annotation update)  
 DE Hypothetical protein Y51H7C.7.  
 GN ORFName=Y51H7C.7;  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Peioderinae; Caenorhabditis.  
 OC NCBI\_Taxid=6239;  
 RX MEDLINE=9069613; PubMed=9851916;  
 RA Wilson R.;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Bradshaw-Cordum H., Dubuque T.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Waterston R.H.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Waterston R.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Waterston R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Waterston R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Waterston R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Waterston R.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Waterston R.;  
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Waterston R.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

RN [11]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Waterston R.;  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Waterston R.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Waterston R.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscol N2;  
 RA Wilson R.;  
 RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC024805; AAK3933.1; -.  
 DR WormPep; Y51H7C.7; C822363.  
 KW Hypothetical protein.  
 SQ SEQUENCE 133 AA; 14852 MW; 1F92FB287391CBF CRC64;

Query Match 18.0%; Score 128.5; DB 2; Length 133;  
 Best Local Similarity 29.3%; Pred. No. 0.0013;  
 Matches 36; Conservative 29; Mismatches 47; Indels 11; Gaps 3;

QY 16 EYLAFENSIGAVDEMKTWVSRLQKLPLEQAKVDVSAVYTLNSMPVYLTQ 75  
 DB 20 ELITKEDAVEEDVGEKHFERSAHM-----ALVDYTMSPFLMSLMAVQATKG 70  
 QY 76 VN-PKCHPVKQLELRIVYNNRVKSEITDKKKAGKLDGASRFVKAWE-PSKNAKV 133  
 DB 71 CGADKDDLLDILAFRTKMTADMKELINLRODAVRINKQAAANVRNALWEPQEGSSKK 130

QY 134 ANK 136  
 DB 131 AAK 133

RESULT 15  
 Q9VXL4 PRELIMINARY; PRT; 159 AA.  
 ID Q9VXL4; 01-MAY-2000 (TReMBLrel. 13, Created)  
 AC Q9VXL4; 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
 DE CG8928-PA.  
 GN ORFName=CG8928;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adame M.D., Celiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champs M., Pfeiffer B.D.,  
 RA Van K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mayhew A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using *frame\_plus\_p2n* model

Run on: January 5, 2005, 12:51:43 : Search time 2500 Seconds  
(without alignments)  
2055.201 Million cell updates/sec

Title: US-09-701-618a-2  
Perfect score: 715  
Sequence: 1 MAGEEINEDYPRVIEHYLSA.....LMPEKSNASKVANKSKSKS 141

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues  
Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+*p2n*.model -DB=xlp  
-Q/cgn2.1/USPTO.spool.p/US09701618/runat.05012005.085839.24447/app\_query.fasta\_1.654  
-DB=EST -QPM=faaap -SUFFIX=xrc -MINMATCH=0.1 -LOOPEXT=0 -LIST=45  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPR=pcp -NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701618\_@CGN\_1.1.6628\_@runat.05012005.085839.24447 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:.\*  
1: gb\_est1:.\*  
2: gb\_est2:.\*  
3: gb\_est3:.\*  
4: gb\_est4:.\*  
5: gb\_est5:.\*  
6: gb\_est6:.\*  
7: gb\_est7:.\*  
8: gb\_est8:.\*  
9: gb\_est9:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	497	6	CA867549 1r29g10.y
2	715	100.0	498	6	CB160277 K-BST0219
3	715	100.0	521	5	BO636280 h066911.y
4	715	100.0	521	6	CD101786 AGENCOURT
5	715	100.0	524	7	CK430831 CJ54d12.y
6	715	100.0	529	7	CF455458 AGENCOURT
7	715	100.0	563	4	BG496371 602541062
8	715	100.0	653	4	BG677242 602623944
9	715	100.0	675	7	CK001081 AGENCOURT

10	715	100.0	738	5	BUS61215 AGENCOURT
11	715	100.0	756	4	BG697814 602661015
12	715	100.0	771	5	BUS928625 AGENCOURT
13	715	100.0	789	6	CB998397 AGENCOURT
14	715	100.0	813	5	BUS68056 AGENCOURT
15	715	100.0	838	5	BUS68457 AGENCOURT
16	715	100.0	872	2	BE886831 601507804
17	715	100.0	963	6	CD515525 AGENCOURT
18	715	100.0	979	5	BQ219896 AGENCOURT
19	710	99.3	469	1	AA424253 zv81e12.r
20	710	99.3	555	4	BG283692 602407460
21	710	99.3	647	6	CD688633 EST5155.h
22	710	99.3	656	6	CB138723 K-BST0191
23	710	99.3	788	1	AV763046 AV763046
24	710	99.3	982	5	BQ434640 AGENCOURT
25	710	99.3	988	4	BM923929 AGENCOURT
26	710	99.3	1014	4	BM451300 AGENCOURT
27	708	99.0	613	4	BG533158 602580684
28	707	98.9	773	6	CB231101 AGENCOURT
29	707	98.9	781	6	CB230406 AGENCOURT
30	707	98.9	794	6	CB309384 AGENCOURT
31	707	98.9	805	6	CB312445 AGENCOURT
32	707	98.9	915	7	CN645648 ILLUMIGEN
33	706	98.7	470	4	BM129833 1F23906.y
34	706	98.7	602	1	AV685739 AV685739
35	705	98.7	554	1	AV686813 AV686813
36	705	98.6	542	2	BF667890 602122126
37	704	98.5	807	5	BUS85132 AGENCOURT
38	702	98.2	577	4	BG502972 602550641
39	702	98.2	779	7	CF596623 AGENCOURT
40	702	98.2	787	6	CD000617 AGENCOURT
41	700.5	98.0	784	2	BF218466 601881839
42	700	97.9	464	7	CN342946 170005322
43	699	97.8	430	1	AI935348 w082d07.x
44	698	97.6	609	4	BG501950 602548269
45	698	97.6	667	4	BG613266 602641127

## ALIGNMENTS

RESULT 1  
CA867549 497 bp mRNA linear EST 20-DEC-2002  
1r29g10.y1 HR85 16let Homo sapiens cDNA clone IMAGE:6546812 5'  
similar to TR:Q13901 Q13901 CID DNA-BINDING PROTEIN. ; mRNA  
sequence.

ACCESSION  
CA867549  
VERSION  
CA867549.1 GI:27319098  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 497)

AUTHORS  
Melton, D., Brown, J., Keny, G., Permutt, A., Lee, C., Kaestner, K.,  
Iemishka, I., Seearce, M., Bretcell, J., Girdwohl, G., Clifton, S.,  
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blustein, A.,  
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,  
Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarisvill, R.,  
Williams, T., Jackson, Y., and Bowers, Y.  
Endocrine Pancreas Consortium

TITLE  
Unpublished (2000)  
CONTACT  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138

TELEPHONE  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@molb.harvard.edu  
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Dr. Hiroshi Inoue



QY 141 Ser 141  
 DB 456 AGT 458

RESULT 3  
 LOCUS B0636280  
 DEFINITION B0636280 521 bp mRNA linear EST 15-JUL-2002  
 Homo sapiens cDNA (Un-normalized, unamplified): hd/he  
 Homo sapiens cDNA clone hd06g11 5', mRNA sequence.

ACCESSION B0636280  
 VERSION B0636280.1 GI:21760739  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 521)  
 Wistow, G., Bernstein, S.L., Wyatt, M.K., Ray, S., Behal, A., Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K. A. Expressed sequence tag analysis of human retina for the NEIBank Project: Rebinding, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts  
 Mol. Vis. 8 (4), 196-204 (2002)  
 M01. Vis. 8 (4), 196-204 (2002)  
 MEDLINE 22103461  
 PUBMED 12107411  
 COMMENT Contact: Wistow G  
 Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: gtwem@helix.nih.gov  
 Plate: 06 row: 9 column: 11  
 Seg primer: M13RP1 reverse primer (ABI).  
 Location/Qualifiers  
 1..521  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="hd06g11"  
 /tissue\_type="Retina"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /clone\_lib="Human Retina cDNA (Un-normalized, unamplified): hd/he"  
 /note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's Instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor  
 15'-pGACTAGTCTAGATCGCGAGCGGCCCGC(7)15-3'. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:

Pred. No.:	Length:	521
Score:	715.00	Matches: 141
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	100.00%	Indels: 0
DB:	5	Gaps: 0

US-09-701-618A-2 (1-141) x B0636280 (1-521)

QY 1 Meta[AG]GluGluLeuAenGluAapTyrProValGluLeuLeuGluTyrLeuSerAla 20

DB 59 ATGCAGGTGAAGAAATTATGAAGACTATCCAGTGAATTCACGAGTATTTGTGACGC 118  
 QY 21 PheGluAsnSer1Ieg1AlaValAapGluMetLeuLysThreMetSerValSerArg 40  
 DB 119 TTTCAGAAATTCATTGGTCTGTGATGAGATCTGAAAGACATGATGCTGTTTCYAGA 178  
 QY 41 AenGluLeuLeuGluGlnLysLeuAapProLeuGluGlnAlaValAapLeuValSerAla 60  
 DB 179 AATGAGTTGTGACGAAGTTGATCCATGAAACGAAGGATTTGGTTTCTGCA 238  
 QY 61 TyrThrLeuAsnSerMetPheTyrValTyrLeuAlaThrGlnGlnValAapProLysGlu 80  
 DB 239 TACACATTAATTCATGATGTTTGGTTTATTTGGCAACCAAGAGTATTCCTTAAGAA 298  
 QY 81 HisProValLysGlnGluLeuGluValArg1LeaValTyrMetAsnArgValLysGlu11e 100  
 DB 299 CATCCAGTAAACAGGAATTCAGAAATCAGAGTATATATGAAACAGATCAAGAAATA 358  
 QY 101 ThrAapLysLysLysValAag1LysLysLeuAapArgG1ValAlaSerArgPheValLysAsn 120  
 DB 359 ACAGACAGAAAGAAAGCTGGCAAGCTGGACAGAGTGACCTTCAGATTTGTAAAAAT 418  
 QY 121 AlALeuTrogLupProLysSerLysAsnAlaSerLysValAlaAsnLysG1LysSerLys 140  
 DB 419 GCCCTCTGGAAACCAATTCGAAATTCATCAAAATTCCTCAATTAAGAAAAAGTAA 478  
 QY 141 Ser 141  
 DB 479 AGT 481

RESULT 4  
 CD101786 521 bp mRNA linear EST 15-MAY-2003  
 LOCUS AGENCOURT\_13980951 NIH MGC\_187 Homo sapiens cDNA clone  
 DEFINITION IMAGE:30373459 5', mRNA sequence.  
 CD101786  
 ACCESSION CD101786.1 GI:30754960  
 VERSION EST.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 521)  
 NIH-MGC http://mgc.nci.nih.gov/  
 Unpublished (1999)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cga@pdx-mail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palokovits  
 cDNA Library Preparation: CLOMTECH Laboratories, Inc.  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: NDCM167 row: e column: 20  
 High quality sequence stop: 521.  
 Location/Qualifiers  
 1..521  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30373459"  
 /lab\_host="DH10B (TI phage-resistant)"  
 /clone\_lib="NIH MGC 187"  
 /note="Organ: Blood\_vessels - aorta, basilar and artery; Vector: pDNR-LIB; Site:1: SfiI (ggccatcatggcc); Site:2: SfiI (ggcgcgtggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCCATTAATGACC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGGCCGCGCCAGTGT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb



KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE NIH-MGC  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Dr. James R. Lupski  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDAMS82 row: 1 column: 14  
High quality sequence stop: 529.  
Location/Qualifiers  
1. 529  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30516205"  
/issue\_type="Peripheral Nervous system"  
/lab\_host="DH10B (TI phage-resistant)"  
/clone\_lib="Lupski anterior horn"  
/note="Vector: pCMV-SPORT6.1; Site\_1: EcoRV (destroyed);  
Site\_2: NotI; Library is oligo-dT primed and directionally  
cloned (EcoRV site is destroyed upon cloning). Average  
insert size 2.1 kb. Library was constructed by Invitrogen  
and donated by J. Lupski, M.D./Ph.D. (Baylor College of  
Medicine)."  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,73e-80 Length: 529  
Score: 715.00 Matches: 141  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-09-701-618a-2 (1-141) x CF455458 (1-529)  
QY 1 MetAlG1yG1uG1u1leAenG1uAePTyPrProva1G1u1leH1eg1uTyLeuSer1a 20  
DB 43 ATGGCAGGTCGAAAGAAATTATGAAGACTATCCAGTGAATTCACGAGTATTGTGACGC 102  
QY 21 PhG1uAenSer1leg1yAlaValaAeG1uMeLeu1yThMetSetSer1aSer1a 40  
DB 103 TTGAGAAATTCATGTGCTGTGATGAGATGCTGAAGACCATGATGCTGTTCTAGA 162  
QY 41 AenG1uLeu1eug1n1yS1eAeAPProLeuG1u1n1a1eYValAeP1eUValSer1a 60  
DB 163 AATGAGTTGTGCAAGATTGCATCCATTCGAAACAGCAAAAGTGGTTTGTGCA 222  
QY 61 TTTThLeuAenSerMetPheTPValTYrLeuA1aThG1nG1yAlaAenProLYeG1u 80  
DB 223 TACACATTAATTCATGTTTGGTTTATTTGGCAACCAAGAGATTATCTTAAGAA 282  
QY 81 H1eAProValYeG1nG1u1eug1n1yAeG1uAeG1u1eAeG1u1eAeG1u1e 100  
DB 283 CATCCAGTAAACAGAAATTCGAAAGATCAGATATATATGAACAGATCAAGAAATA 342  
QY 101 Th1eAP1eY1eY1eY1aG1yLYeLYeAeAPARG1yAla1a1eSerARPhValLYeAen 120  
DB 343 ACAGACAGAAAAAGGCTGGCAAGCTGGACAGAGTGACCTTCAAGATTGTAATAAT 402

QY 121 AlAeuTPG1uPProLYeSer1eAeA1aSer1yVal1aAen1yG1yLeuSer1y 140  
DB 403 GCCCTCGGAAACCAATATGCAAAATGATCAAAAGTTGCAATTAAGAAAAAATA 462  
QY 141 Ser 141  
DB 463 AGT 465  
RESULT 7  
LOCUS BG496371  
DEFINITION 602541062P1 NIH\_MGC\_59 Homo sapiens cDNA clone IMAGE:4672262 5',  
mRNA sequence.  
ACCESSION BG496371 GI:13457888  
VERSION BG496371.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE NIH-MGC  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM488 row: k column: 15  
High quality sequence stop: 536.  
Location/Qualifiers  
1. 563  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4672262"  
/issue\_type="mucocpidermoid carcinoma"  
/lab\_host="DH10B (TI phage-resistant)"  
/clone\_lib="NIH\_MGC\_59"  
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggcgccctggcc); Site\_2: SfiI (ggccatcatggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CAGGCGCTTATGGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGGAGCGGCGGCGCATG-3' (30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
Library."  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,88e-80 Length: 563  
Score: 715.00 Matches: 141  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-09-701-618a-2 (1-141) x BG496371 (1-563)  
QY 1 MetAlG1yG1uG1u1leAenG1uAePTyPrProva1G1u1leH1eg1uTyLeuSer1a 20  
DB 75 ATGGCAGGTCGAAAGAAATTATGAAGACTATCCAGTGAATTCACGAGTATTGTGACGC 134

QY 21 PheGluAnsSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40  
 Db 135 TTTGAGATTCATTGGTGGCTGTGATGAGATGCTGAAGCCATGATGCTGTTCTAG 194  
 QY 41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla 60  
 Db 195 AATGAGTTGTCGAGAGTTGATCCACTTGACACAGCAAGCAAGATTTGGTTCTGCA 254  
 QY 61 TyrThrLeuAnsSerMetPheTTPValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80  
 Db 255 TACCATTTAAATTCATGTTTGGTTTATTTGGCAACCCAGAGATTAATCCTAAGAA 314  
 QY 81 HisProValLysGlnGlnLeuGlnArgIleArgValTyrMetAsnArgValLysGluLe 100  
 Db 315 CATCCAGTAAACAGGATTTGGAAGATCAGATTAATGAAACAGATCAAGAAATA 374  
 QY 101 ThrAspLysLysLysAlaGlyLysLeuAspArgIleAlaLysArgPheValLysAsn 120  
 Db 375 ACAGACAGAAAGAGCTGCAAGCTGACAGAGGTGCAAGCTTCAAGATTTGTAAATAAT 434  
 QY 121 AlaLeuTyrGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140  
 Db 435 GCCCTCGGGAACCAATCGAATAATGCATCAAAAGTTGCCAATTAAGAAAAAGTAA 494  
 QY 141 Ser 141  
 Db 495 AGT 497

RESULT 8  
 BG677242 653 bp mRNA linear EST 01-MAY-2001  
 LOCUS 602623944f1 NCI\_CGAP\_Skn4 Homo sapiens cDNA clone IMAGE:4748963 5',  
 DEFINITION mRNA sequence.  
 BG677242  
 VERSION BG677242.1 GI:13908639  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Mammalia; Eutheria; Primates; Carnivora; Hominiidae; Homo.  
 TITLE NIH-MGC http://mgi.nci.nih.gov/.  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: James Cleaver, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LHM10601 row: g column: 12  
 High quality sequence stop: 649.  
 Location/Qualifiers

FEATURES  
 source  
 1..653  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4748963"  
 /tissue\_type="squamous cell carcinoma"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NCI CGAP Skn4"  
 /note="Organ: skin; Vector: pCMV-SPORT6; Site\_1: NCI;  
 Site\_2: Salt; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.5kb. Library constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 2, 28e-80 Length: 653  
 Score: 715.00 Matches: 141

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0

US-09-701-618a-2 (1-141) x BG677242 (1-653)

QY 1 MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20  
 Db 36 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTGAATTAATTCACAGATTTGTCAGCG 95  
 QY 21 PheGluAnsSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40  
 Db 96 TTTGAGATTCATTGGTGTGCTGTGAGATGAGATCTGAAGCCATGATGCTGTTCTAAG 155  
 QY 41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla 60  
 Db 156 AATGAGTTGTCGAGAGTTGATCCACTTGAAACAGCAAAAGTGATTTGGTTCTGCA 215  
 QY 61 TyrThrLeuAnsSerMetPheTTPValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80  
 Db 216 TACCATTTAAATTCATGTTTGGTTTATTTGGCAACCCAGAGATTATATCCTAAGAA 275  
 QY 81 HisProValLysGlnGlnLeuGlnArgIleArgValTyrMetAsnArgValLysGluLe 100  
 Db 276 CATCCAGTAAACAGGATTTGGAAGATCAGATTAATGAAACAGATCAAGAAATA 335  
 QY 101 ThrAspLysLysLysAlaGlyLysLeuAspArgIleAlaLysArgPheValLysAsn 120  
 Db 336 ACAGACAGAAAGAGCTGCAAGCTGACAGAGGTGCAAGCTTCAAGATTTGTAAATAAT 395  
 QY 121 AlaLeuTyrGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140  
 Db 396 GCCCTCGGGAACCAATCGAATAATGCATCAAAAGTTGCCAATTAAGAAAAAGTAA 455  
 QY 141 Ser 141  
 Db 456 AGT 458

RESULT 9  
 CK001081 675 bp mRNA linear EST 26-NOV-2003  
 LOCUS AGENCOURT\_16390846 NIH\_MGC\_227 Homo sapiens cDNA clone  
 DEFINITION IMAGE:30718850 5', mRNA sequence.  
 CK001081  
 VERSION CK001081.1 GI:38527115  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Carnivora; Hominiidae; Homo.  
 JOURNAL NIH-MGC http://mgi.nci.nih.gov/.  
 COMMENT Unpublished (1999)  
 Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Professor Miklos Palkovits  
 cDNA Library Preparation: Michael Brownstein / Ted Urdin  
 Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: NDCM262 row: m column: 03  
 High quality sequence stop: 545.  
 Location/Qualifiers

FEATURES  
 source  
 1..675  
 /organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_image="30718850"
/tissue_type="Bulk tissue from Human Spinal cord"
/lab_host="DH10B Tonk"
/clone_idb="N1H_MGC_227"
/notes="Organ: brain/CNS; Vector: pDNR-LiB; Site_1: Sfil (directional); Site_2: Sfil (directional); Library is oligo-dt primed and directionally cloned; 5' and 3' adaptors were used in cloning as follows: 5'-BAGCGTGTCTATCAACGACGAGATGTCATTACGCGCGG-3' 5'-ATTCTAGAGGCGCGAGCGCGCCACATG-d(T)30n-1N-3. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected for >0.5kb with an average insert size of 1.3kb Library created in the laboratory of Jonathan Kuo and Ted Udell."

```

ORIGIN	
Alignment Scores:	
Pred. No.:	2,38e-80
Score:	715.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	7
US-09-701-618A-2 (1-141) x CK001081 (1-675)	
Length:	675
Matches:	141
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

Oy		MetAlaGlyIugIuIlleAsnGlnuSpTyrProValGluIleHisGluTyrlenSerla	20
Dd		66 ATGGCAGGTGAAGAAATTATTAAGACACTATCCAGTAGAATAATTCACGAGTATTGCAGCG	125
Oy		21 PheGluAnSerIleGlyAlaValAspGluMetLeuYsrThMetMetSerValSerArg	40
Dd		126 TTGGAGAATTCATGCTGCTGTGGATGAGATGCTGAAGACCATCATGTCTGTTTCTAGA	185
Oy		41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerLa	60
Dd		186 AATAGATTGTTGCAGAAAGTTGATTCACCTTGAAACAAGAAAAGTGATTTGGTTTCTCA	245
Oy		61 TyrThrLeuAnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProbyGlu	80
Dd		246 TACACATTAAATTCATGTTTTGGGTTTATTGGTGCACCAGAGGTTAACTTAAGAA	305
Oy		81 HisProVallysgInuLeuGluuArglleaArgValTyrMetAsnArgVallysgIuile	100
Dd		306 CATCCAGTAAAAACGAGAAATTGGAAGAACATGATATATGAACAGAGTCACGGAATA	365
Oy		101 ThrAspLysLysLysAlaGlyLysLeuAspArgGlyAlaAserArgPheValIyaSn	120
Dd		366 ACAAGCAAGAAAAGGCTGGCAGCTGACAGAGGTGACAGCTTCAAGATTGTAAAAAAT	425
Oy		121 AlaLeuTrpGluProLysSerLysAsnLaseLysValAlaAsnLysGlyLysSerLys	140
Dd		426 GCCCTCTGGGAACCAAATGCCAAAAATGATCAAAAAGTTGCCAATPAAGAAAAAGTAA	485
Oy		141 Ser 141	
Dd		486 AGT 488	
RESULT 10			
BUS61215			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			

**AUTHORS** NIH-MGC <http://mgc.ncl.nih.gov/>.  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC).  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph.D.

Email: [cgabds-rc@mail.nih.gov](mailto:cgabds-rc@mail.nih.gov)  
Tissue Procurement: CLONTECH  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1C6W2814 Row: 0 Column: 11  
High quality sequence stop: 560.  
Location/Qualifiers

FEATURES	SOURCE
Location/Qualifiers	1. -738
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="IMAGE:6592163"
	/lab_host="DH10B (T1 phage-resistant)"
	/clone_1bp="NH1_MGC_82"
	/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgccctcgccg); Site 2: SfiI (ggccatctcgcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCGCCATTATGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGCGCGCCGACATG-dt(30)-BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

	ORIGIN		
	Alignment Scores:		
Pred. No.:	2.67e-80	Length:	738
Score:	715.00	Matches:	141
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0
	US-09-701-618A-2 (1-141) x BU561215 (1-738)		
OY	1 MetAlAGlVGluGluIleLeAnGLuaSPfyrProValGIurIEHISglUYrLEusSerAla 20		
Dd	46 ATGGCAGGTGAAGAATTATTAAGACTCATTCAGTAGAAATCCACGAGTATTCTCACGC 105S		
OY	21 PheGlUaaSeRIeglyVAlavaLapGIumetLeubYrThwmeCteServalSerArg 40		
Dd	106 TTTGAACAATTCCATTGGTGCTGTGTGATAGATGCGCAAGAACCATGATGTCGTTCCTAGA 165S		
OY	41 AenGiLUeuLeUGlnLYseLuasPPrOLEugInUnlaLySVAlaIpLeUVaSeraLa 60		
Dd	166 AAATAATTCTTCACAAGTTGGATCCACTTAAACAAGCAAAGTGGATTGGTTTTCTCGCA 225S		
OY	61 TYrTheAuSnSemerPhetrOvalYlrEuenLAthrIngInyVALaanPROlySGlu 80		
Dd	226 TACACATTAAATTCATgITTTGGCTTTATTtTGcAcCCnAgagITTAAtTCctPAggAA 285S		
OY	81 HIsprOVAlyLGlngluLeUGluNrgIlleaRgaLYrmetASnaRgaVLysGIUIle 100Q		
Dd	286 CATCCAGTAAAACAAGAAATTGGAAAAAGATTCAGATATATATGAACAGATCCAAGAAATA 345S		
OY	101 THraSpLVSLyALgILyLSLeuasPRgilyalaaseRaRGHevaLLySaBn 120Q		
Dd	346 ACAGACMAAAAAAGCTGCCAAGTCGACGAGGTGCAGCTTCAGATTGTAAAAAAT 405S		
OY	121 AlaleUTrGUlpROlysSerLyAsmIAserLyVaLIlaanULYSgyLYVSerLYb 140Q		
Dd	406 GCCCTTGGAACC AAAATCGAAAAATGCACTAAAGTTC CAATTAAAGAAAAAGTAA 465S		
OY	141 Ser 141		



Db 466 AGT 468

RESULT 12  
LOCUS BG697814

DEFINITION BG697814 756 bp mRNA linear EST 07-MAY-2001  
60266101F1 NCI\_CGAP\_Skn3 Homo sapiens cDNA clone IMAGE:4804394 5',  
mRNA sequence.

ACCESSION BG697814

VERSION BG697814.1 GI:13964449

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 756)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

AUTHORS

TITLE

JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: James Cleaver, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
Sequencing by: Incyte Genomics, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Place: LLNL0701 row: m column: 03  
High quality sequence stop: 683.

FEATURES

source

1..756

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4804394"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NCI\_CGAP\_Skn3"

/note="Organ: Skin; Vector: PCMV-SPORT6; Site 1: NciI;  
Site 2: SalI; Cloned unidirectionally. Primer: oligo dT.  
Average insert size 1.5kb. Library constructed by Life  
Technologies. Note: this is a NCI\_CGAP library."

## ORIGIN

## Alignment Scores:

Pred. No.: 2.76e-80 Length: 756

Score: 715.00 Matches: 141

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 4 Gaps: 0

US-09-701-618a-2 (1-141) x BG697814 (1-756)

QY 1 MetAaGlyGluGluIleAsnGluAspTYrProValGluIleHisGluTYrLeuSerAla 20

Db 70 ATGCAAGTGAAGAAATTAATGAAGACTATCCAGTAAATTCACAGATATTTGTCAAGC 129

QY 21 PheGluSerIleGlyAlaValAspGluMetIleuIysThrMetSerValSerArg 40

Db 130 TTGAGAAATTCATGCTGCTGCGATGAGATGCTGAAGACCATGATGCTGTTCTTAA 189

QY 41 AsnGluLeuLeuGlnIlySleuAspProLeuGluGluAlaIlyValAspLeuValSerAla 60

Db 190 AATGAGTGTGTCAGAAAGTTGCATCCATTCAGACAGCAAAAGTGAATTTGCTTCGCA 249

QY 61 TYrThrLeuAsnSerMetPheTYrValTYrLeuAlaThrGlnGlyValAsnProIysGlu 80

Db 250 TACACATTAATTAATCAATGTTTGGGTTTATTGGCAACCCAGAGCTTAATCTTAAGAA 309

QY 81 HisProValIysGlnGluLeuGluArgGlyLeuArgValTYrMetAsnArgValIysGluIle 100

Db 310 CATCCAGTAAACAGAAATTTGGAAAGATCAAGATATATATATGAACAGAGTCAAGAAATA 369

QY 101 ThrAspIlySlyAlaGlyIlyLeuLeuAspArgGlyAlaAlaSerArgPheValIlyAsn 120

Db 370 ACAGACAGAAAGAAAGGCTGGCAAGCTGACAGAGTGCAGCTTCACAGATTTGTAAGAAAT 429

QY 121 AlaLeuTYrGluProIysSerIlyAsnAlaSerIlyValAlaAsnIysGlyIlySerIlyS 140

Db 430 GCCCTCTCGAAGCAACCAATGCAAAATGCAATCAAAAGTTGCCATTAAGAGAAAGTAAA 489

QY 141 Ser 141

Db 490 AGT 492

RESULT 12

LOCUS BU928625

DEFINITION AGENCOURT\_10421966 NIH\_MGC\_57 Homo sapiens cDNA clone IMAGE:6654513  
5', mRNA sequence.

ACCESSION BU928625

VERSION BU928625.1 GI:24117355

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 771)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

AUTHORS

TITLE

JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Clontech Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Place: LNCM2906 row: e column: 09  
High quality sequence stop: 540.

## FEATURES

source

1..771

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6654513"

/tissue\_type="gliblastoma"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NIH\_MGC 57"

/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site 1:  
SfiI (ggccgcctcgcc); Site 2: SfiI (ggccattatggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor  
sequence: 5'-ATCTAGAGGCGCAGGCGGCGCCAGT-dT(30)-BN-3'  
(where B = A, C, G or G and N = A, C, G, or T). Average  
insert size 1.55 kb (range 0.9-4.0 kb). 12/15 clones  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."

## ORIGIN

## Alignment Scores:

Pred. No.: 2.83e-80 Length: 771

Score: 715.00 Matches: 141

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 5 Gaps: 0

US-09-701-618a-2 (1-141) x BU928625 (1-771)

QY 1 MetAaGlyGluGluIleAsnGluAspTYrProValGluIleHisGluTYrLeuSerAla 20





## ORIGIN

**Alignment Scores:**

US-09-701-618A-2 (1-141) x BU568056 (1-813)

RESULT	15
BUS68457	
LOCUS	
DEFINITION	BUS68457 838 bp mRNA linear EST 16-SEP-2002
ACCESSION	S <sub>1</sub> , cRNA sequence.
VERSION	BUS68457
KEYWORDS	BUS68457.1 GI:22918769 EST. <i>Homo sapiens</i> (human)
SOURCE	

## ORGANISM

**COMMENT**

## FEATURES

Source

1.838

## ORIGIN

US-09-701-618A-2 (1-141) X BU568457 (1-838)

27	1	MECAAGLYLGLUGLULLEANGLSAPPLYProValGlu1LeH:GSLTyLeuSerLa	20
	78	ATGACAGGTAAAGAAATTAATGAAGACTATCCAGTAAATTAACGATATTTGTCAAGC	13
Db			
Oy	21	PheGluSnsSerLllegLyAlaValAspGluMecLeuLyStrMetMetSerValSerArg	40
Db	138	TTTGAGAAATTCATTTGTGTGCTGTGATGAGATGCAAGCAACCAATGATCTGTGTTCTAGA	197
Oy	41	ASnGluLeuLeuGluLnuLyLeuAspProLeuGluGlnAlaLyValAspLeuValSerLa	60
Db	198	AATGAGTGTGTGACGAAGTTGGATCCACTTGSAACAAGAAAGTGATTTGGTTTCTGCG	257
Oy	61	TyrThrLeuAsnSerMetPheTyrAlaTyrtLeuAlaPbGrngLyValaAsnProLySGLu	80
Db	258	TACACATTAATTCATATGTTTGGGTTTATTTGGCAACCCAGGAGTTAATCCTTAAGGA	317
Oy	81	HisProValLyGngGlnGluLeuGluLnuArgLLeaArgValTyrtMetAsnArgValLySGLuile	100
Db	318	CATCCAGTAAACACAGAAATTGAAAGATCAGATATATATATGAACAAGATCAACGAAAT	377
Oy	101	ThrAspLyLyValAlaGlyLyLeuAspArgGlyAlaAlaSerArgPheValLyAsn	120
Db	378	ACAGCAACAGAAAAGAGCTGGCAACCTGGACAGAGTCAAGCTTCAGAGATTGTAAAAAAAT	437
Oy	121	AlaLeuTyrGluProLySerLyAsnAlaSerLyValAlaAsnLySGLyLySerLyS	140

Db	438	GCCTCTGGACCAAAATCGAAAAATGCATCAAAAGTTGCCATTAAGGAAAACTAAA	497
Qy	141	Ser 141	
Db	498	AGT 500	

Search completed: January 5, 2005, 16:10:16  
Job time : 2510 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OW protein - nucleic search, using frimae\_pjn model

Run on: January 5, 2005, 08:58:43 ; Search time 320.5 Seconds  
(without alignments)  
2309.418 Million cell updates/sec

Title: US-09-701-618a-2  
Perfect score: 715  
Sequence: 1 MAGEBINEDPVEIHETLSA.....LMPEKSNKSNKSNKSKS 141

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 413486 seqs, 2624710521 residues  
Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frimae\_pjn\_model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool\_p/US09701618/runat\_05012005\_085838\_24427/app\_query.fasta\_1.654  
-DB=N\_Geneseq\_23Sep04 -OPMT=faetap -SUFFIX=ring -MINMATCH=0.1 -LOOPCTL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosome2 -TRANS=human0.cdl  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pjn -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US09701618\_@CGN\_1\_1\_886\_@runat\_05012005\_085838\_24427 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_23Sep04:.\*  
1: geneseqn1980s:.\*  
2: geneseqn1990s:.\*  
3: geneseqn2000s:.\*  
4: geneseqn2001as:.\*  
5: geneseqn2001bs:.\*  
6: geneseqn2002as:.\*  
7: geneseqn2002bs:.\*  
8: geneseqn2003as:.\*  
9: geneseqn2003bs:.\*  
10: geneseqn2003cs:.\*  
11: geneseqn2003ds:.\*  
12: geneseqn2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	1158	3	AAZ43927 Human CID
2	715	100.0	1172	6	ABL61981 Colon ade
3	715	100.0	1172	6	ABK84717 Human CDN
4	715	100.0	1172	10	ADF81400 Leukaemia
5	715	100.0	1172	10	ADF81399 Leukaemia
6	699	97.8	586	12	ACH68585 Human gen

C	7	671	93.8	426	6	AA143974	AA143974 Mammalian
	8	671	93.8	426	6	AA143973	AA143973 Mammalian
	9	667	93.3	420	6	AD128090	AD128090 Human nuc
C	10	667	93.3	420	6	AD128091	AD128091 Human nuc
	11	654	91.5	426	12	ADJ92841	ADJ92841 Human co-
	12	646	90.3	1038	3	AAZ43928	AAZ43928 Murine CI
C	13	644	90.1	420	6	AA143975	AA143975 Mammalian
	14	644	90.1	420	6	AA143976	AA143976 Mammalian
C	15	519	72.6	380	3	AA143977	AA143977 Human sec
	16	266	37.2	204	12	ACH82285	ACH82285 Human gen
	17	150.5	21.0	991	3	AAC49672	AAC49672 Arabidops
	18	150.5	21.0	1000	3	AAC40266	AAC40266 Arabidops
	19	135	18.9	637	4	ABL14531	ABL14531 Drosophila
C	20	115	16.1	699	12	ADP98823	ADP98823 C. albica
	21	110	15.4	555	10	ACC61644	ACC61644 Gene sequ
	22	110	15.4	555	10	ADK64283	ADK64283 Disease t
	23	106	14.8	65	6	ABN53492	ABN53492 Mouse spl
C	24	103	14.4	2705	6	ABL14530	ABL14530 Drosophila
	25	96	13.4	110000	10	ACF67367_12	ACF67367_12 Continuation (13 o
C	26	95	13.4	182624	10	ACF65379	ACF65379 Phototrab
	27	95	13.3	60	6	ABN34297	ABN34297 Human spl
	28	89	12.4	110000	10	ACF67367_43	ACF67367_43 Continuation (44 o
C	29	89	12.4	110000	10	ACF65388_04	ACF65388_04 Continuation (5 of
C	30	88.5	12.4	770	10	ACF66303	ACF66303 Phototrab
C	31	88.5	12.4	110000	10	ACF67367_12	ACF67367_12 Continuation (13 o
C	32	88.5	12.4	110000	10	ACF67367_13	ACF67367_13 Continuation (14 o
C	33	88.5	12.4	182624	10	ACF65379	ACF65379 Phototrab
C	34	88	12.3	110000	10	ACF67367_40	ACF67367_40 Continuation (41 o
C	35	88	12.3	110000	10	ACF65388_07	ACF65388_07 Continuation (8 of
C	36	86.5	12.1	110000	10	ACF67367_25	ACF67367_25 Continuation (26 o
	37	86.5	12.1	110000	10	ACF65386_3	ACF65386_3 Continuation (4 of
	38	85	11.9	456	10	ADH85189	ADH85189 Enterococ
	39	85	11.9	3885	4	ABL13371	ABL13371 Drosophila
	40	85	11.9	8751	4	ABL13370	ABL13370 Drosophila
	41	83.5	11.7	1767	8	ACA33095	ACA33095 Prokaryot
	42	83.5	11.7	1785	10	ADH83717	ADH83717 Enterococ
	43	83.5	11.7	24601	2	AA131160	AA131160 Enterococ
	44	83.5	11.7	24601	6	AB598955	AB598955 Enterococ
	45	83	11.6	3747	8	ACA48336	ACA48336 Prokaryot

## ALIGNMENTS

RESULT 1	AAZ43927	standard; cDNA; 1158 BP.
ID	AAZ43927	
XX	AAZ43927	
AC	AAZ43927	
XX	17-MAR-2000	(first entry)
DT	17-MAR-2000	
XX	Human CID cDNA.	
XX	CID; human; apoptosis; tumour; gene therapy; treatment; ss.	
KM	CID; human; apoptosis; tumour; gene therapy; treatment; ss.	
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	Key	Location/Qualifiers
FM	CDS	118..543
FT		/*tag= a
FT		/product= "CID"
XX	DE19824811-A1.	
XX	09-DEC-1999.	
PD	09-DEC-1999.	
XX	03-JUN-1998;	98DE-01024811.
PF	03-JUN-1998;	
XX	03-JUN-1998;	98DE-01024811.
PR	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.	
PA	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.	
XX	Rotbarth K, Stammer H, Werner D;	
PI	Rotbarth K, Stammer H, Werner D;	

XX WPI: 2000-063506/06.  
DR P-PSDB; AAY51024.

XX Inducing apoptosis by overexpressing the CID gene, particularly for  
PT treating tumors.

XX Claim 3; Fig 1; 10pp; German.

XX This invention describes a novel method for inducing apoptosis which  
CC comprises overexpressing the CID gene (I). The method is particularly  
CC used to treat tumors and can also be used in gene therapy. The method has  
CC no side effects on normal cells (contrast known methods of inducing  
CC apoptosis such as cytotoxins and radiation), and may be effective on  
CC cells resistant to conventional treatments. Overexpression of (I) is  
CC sufficient itself to induce apoptosis but the effect may be increased  
CC when used in combination with other anti-tumor methods. When cells  
CC transfected with (I) undergo apoptosis, they release factors that kill  
CC neighboring, non-transfected cells (bystander effect). This sequence  
CC encodes the human CID protein described in the method of the invention  
XX

Sequence 1158 BP; 388 A; 155 C; 237 G; 378 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	8,98e-79	Length:	1158
Score:	715.00	Matches:	141
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-701-618A-2 (1-141) \* AAZ43927 (1-1158)

QY 1 MetAaGlyGluGluIleAaNgLuAaPyrProValGluIleHiaGluTyrLeuSerAla 20  
DB 118 ATGCGAGGTGAAGAATTATGAGACATCTCAGTGAATTCACAGATATTGTCAAGC 177  
QY 21 PheGluAaSerIleGlyAlaValAaPgluMetLeuYsrHMetMetSerValSerArg 40  
DB 178 TTGGAATTCATGTTGGTGTGTGATGAGATGCGAAGCATGATGCTGTTCTTCA 237  
QY 41 AaNgLuLeuGluIleAaPgluMetLeuYsrHMetMetSerValSerAla 60  
DB 238 AATGAGTTGTTGCAAGATTGCACTTCAACAGCAAGAGTGGATTGGTTCTGCA 297  
QY 61 TyrThrLeuAaSerMetPheTyrValTyrLeuAlaThrGluGlyValAaPgluSerGlu 80  
DB 298 TACACATTAAATTCATGTTTGGGTTTATTGGCAACCCAGAGATTAACTTCAAGAA 357  
QY 81 HisProValIlyGluGluIleAaPgluMetLeuYsrHMetMetSerValSerAla 100  
DB 358 CATCCAGTAAACAGAGATTGCAAGATCAGATATATATGAAACAGATCAAGGAAATA 417  
QY 101 ThrAspIlySerIlyAaGlyValAaPgluMetLeuYsrHMetMetSerValSerAla 120  
DB 418 ACAAGCAAGAAAGAGCTGGCAAGCTGCAAGAGTGCACACTTCAAGATTGTAAATAAT 477  
QY 121 AlaLeuTrpGluPyrSerIlyAaSerIlyValAaAaGlySerIlySerGlySerGly 140  
DB 478 GCCCTCTGGGAACCAAAATGCAAAATGATCAAAAGTGCCTAAATAAGGAAATAATA 537  
QY 141 Ser 141  
DB 538 AGT 540

#### RESULT 2

ABL61981  
ID ABL61981 standard; DNA; 1172 BP.

XX ABL61981;

DT 15-MAY-2002 (first entry)

DE Colon adenocarcinoma related gene sequence SEQ ID NO:318.

XX Human; Cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;  
XX gene; ds.

OS Homo sapiens.

PN WO200194629-A2.

PD 13-DEC-2001.

PF 30-MAY-2001; 2001WO-US010838.

XX 05-JUN-2000; 2000US-0209473P.  
XX 18-SEP-2000; 2000US-0209531P.  
XX 18-SEP-2000; 2000US-0233133P.  
XX 20-SEP-2000; 2000US-0233617P.  
XX 20-SEP-2000; 2000US-0234009P.  
XX 20-SEP-2000; 2000US-0234034P.  
XX 20-SEP-2000; 2000US-0234052P.  
XX 22-SEP-2000; 2000US-0234509P.  
XX 22-SEP-2000; 2000US-0234567P.  
XX 25-SEP-2000; 2000US-0234923P.  
XX 25-SEP-2000; 2000US-0234924P.  
XX 25-SEP-2000; 2000US-0235077P.  
XX 25-SEP-2000; 2000US-0235082P.  
XX 25-SEP-2000; 2000US-0235134P.  
XX 25-SEP-2000; 2000US-0235280P.  
XX 26-SEP-2000; 2000US-0235637P.  
XX 26-SEP-2000; 2000US-0235638P.  
XX 27-SEP-2000; 2000US-0235711P.  
XX 27-SEP-2000; 2000US-0235720P.  
XX 27-SEP-2000; 2000US-0235840P.  
XX 27-SEP-2000; 2000US-0235863P.  
XX 28-SEP-2000; 2000US-0236028P.  
XX 28-SEP-2000; 2000US-0236032P.  
XX 28-SEP-2000; 2000US-0236033P.  
XX 28-SEP-2000; 2000US-0236034P.  
XX 28-SEP-2000; 2000US-0236109P.  
XX 28-SEP-2000; 2000US-0236111P.  
XX 29-SEP-2000; 2000US-0236842P.  
XX 29-SEP-2000; 2000US-0236891P.  
XX 02-OCT-2000; 2000US-0237172P.  
XX 02-OCT-2000; 2000US-0237173P.  
XX 02-OCT-2000; 2000US-0237278P.  
XX 02-OCT-2000; 2000US-0237294P.  
XX 02-OCT-2000; 2000US-0237295P.  
XX 02-OCT-2000; 2000US-0237316P.  
XX 03-OCT-2000; 2000US-0237425P.  
XX 03-OCT-2000; 2000US-0237598P.  
XX 03-OCT-2000; 2000US-0237604P.  
XX 03-OCT-2000; 2000US-0237606P.  
XX 03-OCT-2000; 2000US-0237607P.  
XX 01-NOV-2000; 2000US-0244867P.  
XX 01-NOV-2000; 2000US-0245084P.

PA (AVAL-) AVALON PHARM.

PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S,  
PI Soppet DR, Weaver Z;

DR WPI; 2002-188264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
PT agent to be tested for anti-neoplastic activity, and determining a change  
XX in expression of a gene of a signature gene set.

XX Claim 1; SEQ ID NO 318; 44pp; English.

XX The present invention describes a method (M1) for screening for an anti-  
CC neoplastic agent. The method involves exposing cells to a chemical agent



```
US-09-701-618A-2 (1-141) X ABR84717 (1-1172)
QY      1 Met1AG1yG1uG1u11eAenG1uAepTyrProValG1u11eh1sg1uTyrLeuSerAla 20
Db      118 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAAGAAATTCACGAGTATTTGTCAGCG 177
QY      21 PheG1uAenSer11eg1yAlaValAAspG1uMe1eulYerTherMetSerValSerArg 40
Db      178 TTGAGAAATTCATTTGGTGTGCTGTGATGAGATGCTGAAAGCCAAAGATGCTGTTCTTCA 237
QY      41 AsnG1uLeuLeuG1u1yLeuAepProLeuG1uG1u11a1yValAspLeuValSerAla 60
Db      238 AATGAGTTGTTCAGAGATTCAGATCCATTCAGCAAGCAAAAGTGAATTTGGTTCTGCA 297
QY      61 TyrThrLeuAenSerMetPheTyrVal1TyrLeuAlaThrG1uG1yValAsnProLySG1u 80
Db      298 TACACATTAATTCATTTGAGTTTATTTGGCAACCCAAAGAGATTATCTTAAGGAA 357
QY      81 HisProVal1ySG1uG1uLeuG1uArg11eArgVal1TyrMetAsnArgVal1ySG1u11e 100
Db      358 CATCCAGTAAACAGGAATTCGAAAGATCAAGTATATATGAACAGAGTCACAGAAATA 417
QY      101 ThrAspLyLe1ySG1yAlaG1yLyLeuAepA-rg1yAla1SerArgPheVal1ySG1u 120
Db      418 ACGACAAAGAAAGAGCTGCGCAAGCTGACAGAGGTGCGAGCTTCAAGATTGTAAAAAAT 477
QY      121 AlaLeuTyrG1uProLySer1ySG1yAsnAlaSerLySG1yVal1AAsnLySG1yLySG1y 140
Db      478 GCCCTCTGGGAACCAAAATCGAAAAATGATCAAAAGTTGCCATTAAGAAAAAGTAA 537
QY      141 Ser 141
Db      538 AGT 540

RESULT 4
ADF81400
ID      ADF81400 standard; DNA; 1172 BP.
XX
AC      ADF81400;
XX
DT      26-FEB-2004 (first entry)
XX
DE      Leukaemia-related DNA sequence #1956.
XX
KW      Cytostatic; Gene therapy; leukaemia; ss.
XX
OS      Unidentified.
XX
PN      WO2003039443-A2.
XX
PD      15-MAY-2003.
XX
PE      04-NOV-2002; 2002WO-EP012303.
XX
PR      05-NOV-2001; 2001EP-00126244.
XX
PR      30-APR-2002; 2002EP-00009758.
XX
PA      (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PA      (UYLU-) UNIV LUDWIG MAXIMILIANS.
XX
PA      (HAFEL/) HAFERLACH T.
XX
PA      (SCHO/) SCHOCH C.
XX
PA      (KERN/) KERN W.
XX
PI      Haeferlach T, Schoch C, Kern W, Kohlmann A, Schittiger S, Dugas M,
XX
PI      Ellis R, Brors B, Mezenthaier S;
XX
DR      WPI; 2003-505037/47.
XX
PT      Determining the subtype of leukemia cells and whether a patient sample
XX
PT      contains leukemia cells or other cells, useful for treating leukemia,
XX
PT      comprises determining the expression profile of a group of markers in a
XX
PT      patient sample.
```

```
XX
PS      Disclosure; SEQ ID NO 1956; 293bp; English.
XX
CC      The present invention relates to a method (M1) for determining the
CC      subtype of leukemia cells and whether a patient sample contains
CC      leukemia cells. The method comprises determining the expression profile
CC      of a group of markers in a patient sample. The method is useful for
CC      determining the presence of leukemia cells, its types or subtypes, and
CC      for the preparation of a medicament for treating leukaemia.
XX
SQ      Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 9,13e-79 Length: 1172
Score: 715.00 Matches: 141
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-701-618A-2 (1-141) X ADF81400 (1-1172)
QY      1 Met1AG1yG1uG1u11eAenG1uAepTyrProValG1u11eh1sg1uTyrLeuSerAla 20
Db      118 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAAGAAATTCACGAGTATTTGTCAGCG 177
QY      21 PheG1uAenSer11eg1yAlaValAAspG1uMe1eulYerTherMetSerValSerArg 40
Db      178 TTGAGAAATTCATTTGGTGTGCTGTGATGAGATGCTGAAAGCCAAAGATGCTGTTCTTCA 237
QY      41 AsnG1uLeuLeuG1u1yLeuAepProLeuG1uG1u11a1yValAspLeuValSerAla 60
Db      238 AATGAGTTGTTCAGAGATTCAGATCCATTCAGCAAGCAAAAGTGAATTTGGTTCTGCA 297
QY      61 TyrThrLeuAenSerMetPheTyrVal1TyrLeuAlaThrG1uG1yValAsnProLySG1u 80
Db      298 TACACATTAATTCATTTGAGTTTATTTGGCAACCCAAAGAGATTATCTTAAGGAA 357
QY      81 HisProVal1ySG1uG1uLeuG1uArg11eArgVal1TyrMetAsnArgVal1ySG1u11e 100
Db      358 CATCCAGTAAACAGGAATTCGAAAGATCAAGTATATATGAACAGAGTCACAGAAATA 417
QY      101 ThrAspLyLe1ySG1yAlaG1yLyLeuAepA-rg1yAla1SerArgPheVal1ySG1u 120
Db      418 ACGACAAAGAAAGAGCTGCGCAAGCTGACAGAGGTGCGAGCTTCAAGATTGTAAAAAAT 477
QY      121 AlaLeuTyrG1uProLySer1ySG1yAsnAlaSerLySG1yVal1AAsnLySG1yLySG1y 140
Db      478 GCCCTCTGGGAACCAAAATCGAAAAATGATCAAAAGTTGCCATTAAGAAAAAGTAA 537
QY      141 Ser 141
Db      538 AGT 540

RESULT 5
ADF81399
ID      ADF81399 standard; DNA; 1172 BP.
XX
AC      ADF81399;
XX
DT      26-FEB-2004 (first entry)
XX
DE      Leukaemia-related DNA sequence #1955.
XX
KW      Cytostatic; Gene therapy; leukaemia; ss.
XX
OS      Unidentified.
XX
PN      WO2003039443-A2.
XX
PD      15-MAY-2003.
XX
PE      04-NOV-2002; 2002WO-EP012303.
```



XX 05-NOV-2001; 2001EP-00126244.  
 PR 30-APR-2002; 2002EP-00009758.  
 XX  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
 PA (UJLV-) UNIV LUDWIG MAXIMILIANS.  
 PA (HAF/) HAFERLACH T.  
 PA (SCH/) SCHUCH C.  
 PA (KERN/) KERN W.  
 XX  
 PI Haeferlach T, Schoch C, Kern W, Kohlmann A, Schnitger S, Dugas M,  
 PI Ellis R, Broers B, Mergenthaler S,  
 XX WPI; 2003-505037/47.  
 DR  
 XX  
 PT Determining the subtype of leukemia cells and whether a patient sample  
 PT contains leukemia cells or other cells, useful for treating leukemia,  
 PT comprises determining the expression profile of a group of markers in a  
 PT patient sample.  
 XX  
 PS Disclosure; SEQ ID NO 1955; 293bp; English.  
 XX  
 CC The present invention relates to a method (M1) for determining the  
 CC subtype of leukemia cells and whether a patient sample contains  
 CC leukemia cells. The method comprises determining the expression profile  
 CC of a group of markers in a patient sample. The method is useful for  
 CC determining the presence of leukemia cells, its types or subtypes, and  
 CC for the preparation of a medication for treating leukemia.  
 XX  
 SQ Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 9.13e-79 Length: 1172  
 Score: 715.00 Matches: 141  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0  
 US-09-701-618a-2 (1-141) x ADF81399 (1-1172)  
 QY 1 MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHISGluTyrLeuSerAla 20  
 Db 118 ATGGCAGGTGAAGAAATTTATGAAGACTATCCAGTAGAAATTCGAGATTTTCACGC 177  
 QY 21 PheGluAnsSerIleGlyAlaValAspGluMetLeuIleThrMetMetSerValSerArg 40  
 Db 178 TTTCAGAAATTCATTTGGTGGTGGATGAGATGCTGAAGACCATGATGCTGTTCTAGA 237  
 QY 41 AsnGluLeuLeuGluIleValLeuAspProLeuGluGlnAlaValAspLeuValSerAla 60  
 Db 238 AATGAGATTTGTCGAAAGTTGGATCCCTGGAACAGCAAAAGTGATTTGGTTCTGCA 297  
 QY 61 TyrThrLeuAnsSerMetPheTyrValTyrLeuAlaThrGlnGlyValAspProValGlu 80  
 Db 298 TACCATTAATTAATGTTTGGTTTATTTGGCAACCCAGAGATTATCTTAAGGAA 357  
 QY 81 HisProValIleGlnGluLeuGluArgIleArgValTyrMetAsnArgValIleGluIle 100  
 Db 358 CATCACTAAACAGCAAGATGGAAGATCATGATATATGACAGTCAAGGAAATA 417  
 QY 101 ThrAspIleValIleValAlaGlyValLeuAspArgGlyAlaAlaSerArgPheValIleAsn 120  
 Db 418 ACAGACAAAGAAAGGCTGGCAGCTGACAGAGTCAAGTTCAGATTTGTAATAAAT 477  
 QY 121 AlaLeuTrpGluProIleSerIleValAsnAlaSerIleValIleAsnValGlyValSerIle 140  
 Db 478 GCCCTTCGGGAACCAATTCGAAATGCAATCAAAAGTTGCAATTAAGGAAAGTAATA 537  
 QY 141 Ser 141  
 Db 538 AGT 540

RESULT 6  
 ACH68585/C  
 ID ACH68585 standard; DNA, 586 BP.  
 XX  
 AC ACH68585;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE Human genome derived single exon probe #1780.  
 XX  
 KW Human; probe; ss; gene expression; single exon probe; microarray;  
 KW alternative splicing event; genomic alteration.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003194704-A1.  
 XX  
 PD 16-OCT-2003.  
 XX  
 PF 03-APR-2002; 2002US-00029386.  
 XX  
 PR 03-APR-2002; 2002US-00029386.  
 XX  
 PA (PENN/) PENN S G.  
 PA (RANK/) RANK D R.  
 PA (HANZ/) HANZEL D K.  
 XX  
 PI Penn SG, Rank DR, Hanzel DK;  
 XX  
 DR WPI; 2004-119264/12.  
 XX  
 PT New human genome-derived single exon nucleic acid probes useful for human  
 PT gene expression analysis, for identifying or characterizing alternative  
 PT splicing events, for assessing genomic alterations or as tools for  
 XX surveying tissues.  
 XX  
 PS Claim 15; SEQ ID NO 1780; 80bp; English.  
 XX  
 CC The invention relates to a nucleic acid probe for measuring human gene  
 CC expression, comprising any of the 27,400 fully defined nucleotide  
 CC sequences in the specification, or their complements or fragments, and  
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences  
 CC fully defined in the specification. The probe is a single exon probe that  
 CC hybridizes under high stringency conditions to a nucleic acid molecule  
 CC expressed in human cells or tissues. Also included are a spatially-  
 CC addressable set of single exon nucleic acid probes for measuring human  
 CC gene expression (comprising a plurality of single exon nucleic acid  
 CC probes cited above, where each of the plurality of probes is separately  
 CC and addressably isolatable or amplifiable from the plurality), a single  
 CC exon microarray for measuring human gene expression, a method of  
 CC measuring human gene expression, a vector comprising the single exon  
 CC probe cited above, an ORF-encoded peptide comprising at least 8  
 CC contiguous amino acids of any of the above-mentioned amino acid  
 CC sequences (optionally with conservative amino acid substitutions), an  
 CC isolated antibody that binds specifically to a peptide cited above,  
 CC a customer desiring to measure gene expression, a method of providing  
 CC human gene expression data by subscription, and a computer-readable  
 CC storage medium which contains a database having a plurality of records  
 CC (each record including data on the expression of a single exon probe  
 CC cited above. The probe, methods and apparatus are useful in gene  
 CC expression analysis. The probes may be used as tools for surveying  
 CC tissues to detect the presence of expressed messages that contain their  
 CC specific exon, or in constructing genome-derived single exon microarrays.  
 CC In addition, the probes are used in identifying and characterizing  
 CC alternative splicing events, in detecting and characterizing gross  
 CC alterations in the genomic locus that includes their exon, in assessing  
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
 CC or in expressing the ORF-encoded peptide. The present sequence is a human  
 CC single exon probe of the invention. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?docID=20030194704

PT useful for inhibiting cellular function of cofactor and for treating  
PT metabolic disorders, immunological indications and hormonal dysfunctions.  
PS  
XX Claim 1; Fig 2; 68pp; English.

CC The invention comprises the amino acid and coding sequences of two  
CC mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and  
CC CF8 protein sequences of the invention are useful for screening agents  
CC that are capable of inhibiting the cellular function of cofactor CF7  
CC and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes  
CC involved in cellular functions, such as: regulation of metabolism and  
CC cell homeostasis, cell proliferation and differentiation, pathological  
CC cellular aberrations, or cellular defence mechanisms. The present cDNA  
CC sequence represents the reverse complement of the mammalian nuclear  
CC receptor cofactor CF7 gene sequence  
XX

SQ Sequence 426 BP; 105 A; 99 C; 64 G; 156 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 6.76e-74 Length: 426  
Score: 671.00 Matches: 132  
Percent Similarity: 95.74% Conservative: 3  
Best Local Similarity: 93.62% Mismatches: 6  
Query Match: 93.85% Indels: 0  
DB: Gaps: 0

US-09-701-618A-2 (1-141) x AAL43974 (1-426)

QY 1 MetalaglygluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20  
Db 426 ATGCACACTGAAGAAATTATATGAGGACATTCACATGAAATTCACGATTATTGTTCAGCA 367  
QY 21 PheGluAsnSerIleGlyAlaValAspGluMetLeuIleGlySerMetSerValSerArg 40  
Db 366 TTTTGGCAATTCATTTATATCTGTGATGATGAGATGCGAAGAACATGATGCTGTTCTAGA 307  
QY 41 AsnGluLeuGluGlnIleLeuAspProLeuGluGlnAlaIleValAspLeuValSerAla 60  
Db 306 AATGAGATTGTTGACAGAGTTGACCCACTTAAACAGCAAAAGTCGATTGGTTCTGCA 247  
QY 61 TyrThleuAsnSerMetPheTyrValIleTyrLeuAlaThrGlnIleValAsnProIleGlu 80  
Db 246 TACACATTAATTCATGTTGGTTTATTTTGGCACTCAAGAGGATGATCCTTAAGAA 187  
QY 81 HisProValIleGlnIleGluIleGluIleArgIleArgValTyrMetAsnArgValIleGluIle 100  
Db 186 CATCCAGTAAACACAGAAATTGGAAGAATCGAGTATATATATGAACGAGTCACAGGAATA 127  
QY 101 ThrAspIleValIleValAlaGlyIleLeuAspArgIleValAlaIleSerArgPheValIleAsn 120  
Db 126 ACAGCAAGAAAGAAAGGCTGCGCACTGGACGAGGTGCACCTTCAAGATTGTGAAGAA 67  
QY 121 AlaLeuTyrGluIleProIleSerIleAsnAlaSerIleValAlaIleAsnIleGlyIleSerIle 140  
Db 66 GCCCTCTGGAAACCAAAACCGAATAATATCATCCAAAGTTGCCCATATAAGAGAAAAAGTAA 7  
QY 141 Ser 141  
Db 6 AGT 4

RESULT 8  
ID AAL43973  
XX AAL43973 standard; cDNA; 426 BP.  
XX  
XX AAL43973;  
XX  
DT 27-SEP-2002 (first entry)  
XX  
XX Mammalian nuclear receptor cofactor CF7 coding sequence.  
DE  
XX Mammalian; gene; ss; nuclear receptor cofactor; CF7; CF8;  
XX metabolism regulation; cell homeostasis; cell proliferation;  
XX

XX		cellular defence mechanism.
KX	Mammalia.	
OS		
XX		
FX	Key	Location/Qualifiers
FT	CDS	1..426
FT		/tag= a
FT		/product= "Mammalian CF7 protein"
XX		
FN	WO200242322-A2.	
PD		
XX	30-MAY-2002.	
XX		
PE	21-NOV-2001; 2001WO-BP013548.	
XX		
PR	21-NOV-2000; 2000EP-00125524.	
XX		
PA	(LION-) LION BIOSCIENCE AG.	
XX		
PI	Jackson D, Casari G, Suckow J;	
DR	WPI; 2002-566559/60.	
DR	P-PsDB; AA015405.	
XX		
PT	Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators	
PT	useful for inhibiting cellular function of cofactor and for treating	
PT	metabolic disorders, immunological indications and hormonal dysfunctions.	
PS		
XX	Claim 1; Fig 1; 68pp; English.	
CC		
XX	The invention comprises the amino acid and coding sequences of two	
CC	mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and	
CC	CF8 protein sequences of the invention are useful for screening agents	
CC	that are capable of inhibiting the cellular function of cofactor CF7	
CC	and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes	
CC	involved in cellular functions, such as: regulation of metabolism and	
CC	cell homeostasis, cell proliferation and differentiation, pathological	
CC	cellular aberrations, or cellular defence mechanisms. The present cDNA	
CC	sequence encodes the mammalian nuclear receptor cofactor CF7 protein	
SQ		
XX	Sequence 426 BP; 158 A; 64 C; 99 G; 105 T; 0 U; 0 Other:	
Alignment Scores:		
Pred. No.:	6,76e-74	Length: 426
Score:	671.00	Matches: 132
Percent Similarity:	95.74%	Conservative: 3
Best Local Similarity:	93.62%	Mismatches: 6
Query Match:	93.85%	Indels: 0
DB:	6	Gaps: 0
US-09-701-618A-2 (1-141) x AAL43973 (1-426)		
OY	1 MetAlAGLyGluGlnIleAsnGluApPrProValGIuIHSGluTrLeuSerAla 20	
Db		
	1 ATGGCAGCTGAAGAATTATATGAGCACTATCCACGTAGAAATTCACCATTTTGTACACA 60	
OY	21 PheGIuAnSerIleGlYAlaValAPrGluMetLeuIleYThrMetMetSerValSerArg 40	
Db		
	61 TTTCGGAATTCATTGATGATGCTGTGCATGAGTGACTGAAGAACATGATGCTGTTCTACA 120	
OY	41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla 60	
Db		
	121 AATGAGTTGTTGCCAAGATGTGACCACCTTGAAACAAGCAAAGTGGATTGGTTCTGCA 180	
OY	61 TyrThrLeuAsnSerMetPheTrpValTYrLeuAlaThGlnGlyValAsnPcrlvGlu 80	
Db		
	181 TACACATTTTAATTCAATGTTTTGGGTTTATTTGGCACTCAAGAGAGAAATCTTAAGGAA 240	
OY	81 HisProValYyGInGluLeuGluIuArgIleArgValTYrMetAsnArgValYyGluIle 100	
Db		
	241 CATCCAGTAAACACAGGAATTTGGAAGAATCAAGAGTATATATGACAGAGTCACAGGAATA 300	
OY	101 ThrAspLYbLYbLYbLYbGLyLYbLYbLeuAspArgGlyAlaLAserArgPheValLYbAsn 120	

Db	301	ACGACACAGAAAAAGGCTGCAAGCTGGACAGAGGTCAGCTTCAAGATTGTGAAGAAAT	366
Qy	121	AlaLeuTPGIuProLysSerLysAlaSerLysValAlaAlaSerLysSerLys	140
Db	361	GGCCTCTGGGACCAAAACCGAAAAATATCATCCAAAGTTGCCCATTAAGAAAAAGTAAA	420
Qy	141	Ser 141	
Db	421	AGT 423	
RESULT 9			
ID	AD128090	standard; DNA; 420 BP.	
XX	AC	AD128090;	
XX	DT	22-APR-2004 (first entry)	
XX	DE	Human nuclear receptor cofactor CF6 sense coding strand.	
XX	KW	gene; nuclear receptor cofactor; CF6; cellular function inhibition;	
KW	KM	metabolic disorder; immunological indication; hormonal dysfunction;	
KW	KM	neurosystemic disease; ss.	
XX	OS	Homo sapiens.	
XX	FH	Key	Location/Qualifiers
XX	FT	CDS	1..420
FT			/*tag= a
XX	PN	WO200224728-A2.	
XX	PD	28-MAR-2002.	
XX	PF	17-SEP-2001; 2001MO-EP010744.	
XX	PR	22-SEP-2000; 2000EP-00120722.	
XX	PA	(LION-) LION BIOSCIENCE AG.	
PI		Caesari G, Jackson D;	
XX	DR	WPI; 2002-383179/41.	
XX	DR	P-P5DB; AD128083.	
PT		Polynucleotide and polypeptide of novel nuclear receptor cofactor useful	
PT		for screening drugs regulating cofactor-associated physiological	
XX		responses e.g. hormonal dysfunctions.	
PS		Claim 1; SEQ ID NO 1; 97bp; English.	
XX		The invention relates to an isolated polynucleotide encoding a nuclear	
CC		receptor cofactor, also known as CF6. The polynucleotide or encoded	
CC		protein is useful for construction of multiple nuclear receptor cofactor	
CC		specific sequence alignments, preferably for the construction of protein	
CC		sequence alignments. The protein is useful for screening agents capable	
CC		of inhibiting the cellular function of the cofactor CF6. The	
CC		polynucleotide is useful for making vectors and for transforming cells,	
CC		both of which are ultimately useful for production of the CF6 protein.	
CC		They are also useful as scientific research tools for developing nucleic	
CC		acid probes for determining expression levels of the cofactor gene, e.g.	
CC		to identify diseased or otherwise abnormal states. They are particularly	
CC		useful for diagnostic purposes to e.g., identify deleted or mutant CF6	
CC		genes; or their measure expression. They are useful for developing	
CC		analytical tools such as antisense oligonucleotide for selectively	
CC		inhibiting expression of the cofactor gene to determine physiological	
CC		responses. The protein is useful for screening drugs for agonist and	
CC		antagonist activity, and therefore, for screening for drugs useful in	
CC		regulating physiological responses associated with the cofactors such as	
CC		metabolic disorders, immunological indications, hormonal dysfunction,	
CC		neurosystemic diseases. The proteins are also useful for developing	
CC		antibodies for detection of the proteins. The polynucleotide can be used	



Db 60 GCCCTCTGGAGAACCAACGAAATATACATCCAAAGTGGCCATTAAGAAAAATGAAA 1  
 RESULT 11  
 ADJ92841  
 ID ADJ92841 standard; DNA; 426 BP.  
 XX  
 AC ADJ92841;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Human co-repressor SUN-COR DNA.  
 XX  
 KM Nuclear receptor; therapy; diabetes; hyperlipidaemia; obesity;  
 KM atherosclerosis; human; co-repressor; gene; de.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003228607-A1.  
 XX  
 PD 11-DEC-2003.  
 XX  
 PF 14-APR-2003; 2003US-00414692.  
 XX  
 PR 15-APR-2002; 2002US-0372650P.  
 XX  
 PA (WAGN/) WAGNER B L.  
 PA (SCHU/) SCHULMAN I G.  
 PI Wagner BL, Schulman IG;  
 XX  
 DR WPI; 2004-167207/16.  
 XX  
 PT Identifying compounds that bind to nuclear receptor and exhibit cell type  
 PT specific actions, and useful for treating hyperlipidemia, obesity and  
 PT diabetes.  
 XX  
 PS Disclosure; SEQ ID NO 42; 99pp; English.  
 XX  
 CC The invention relates to screening methods for identifying compounds that  
 CC bind to nuclear receptor and exhibit cell type specific actions. The  
 CC invention relates to modulators having an improved therapeutic profile.  
 CC The method is useful for identifying compounds that bind to a nuclear  
 CC receptor and exhibit cell type specific actions. It is also useful for  
 CC identifying modulators of nuclear receptors that are useful in treating  
 CC diseases e.g., diabetes, hyperlipidemia, obesity, atherosclerosis, etc.  
 CC The present sequence is human co-repressor DNA used to illustrate the  
 CC method of the invention.  
 XX  
 SQ Sequence 426 BP; 150 A; 76 C; 107 G; 93 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 8,74e-72 Length: 426  
 Score: 654.00 Matches: 127  
 Percent Similarity: 96.43% Conservative: 8  
 Best Local Similarity: 90.71% Mismatches: 5  
 Query Match: 12 Gaps: 0  
 DB: 12  
 US-09-701-618a-2 (1-141) x ADJ92841 (1-426)  
 QY 1 MetAlAGlGlnGluLueAaGUAaPTTyrProValGluLueHsGluTyrLueSerAla 20  
 Db 1 ATGGCGAGGTAAAGAAATGAATGAAGATTATCCGTAAATAATTCACGAGTCTTTAAACGCC 60  
 QY 21 PheGluAsnSerIleGlyAlaValAaPGLuMetLueLysThrMetSerValSerArg 40  
 Db 61 CTGAGAGACGCTCCCTGGCTGCTGTGATGACATGCTGAAGACCATGATGCGCTTTCTAGA 120  
 QY 41 AaGluLueLueGlnLysLueAaPProLueGlnGlnAlaLysValAaPLeuValSerAla 60  
 Db 121 AATGAGATTGTGGAAGATTGACCCATTTGGAACAAAGGAGGATTAGTTTGTGCA 180  
 QY 61 TyrThrLueAsnSerMetPheTyrPValTyrLueAlaThrGlnGlyValaAnProLysGlu 80

Db 181 TACACCTTAATTCANGTGTTGGTTATTGGCAACTACAGAGATTATCCAAAGAG 240  
 QY 81 HisProValLysGlnGluLueGluAaGlyIleArgValTyrMetLsnArqValLysGluIle 100  
 Db 241 CATCCAGTGAAGCAGAACTCGAAGAAATCGATCTACATGAACAGAGTTTAAAGAAATA 300  
 QY 101 ThrAspLysLysLysAlaGlyLysLueAaPArgGlyValAlaLsSerArgPheValLysAsn 120  
 Db 301 ACAGACAAAGAAAGGCTGCCAAGCTGACAGAGGCTCTTCGACATTGTCAAGAAC 360  
 QY 121 AlaLueTyrPGLuProLysSerLysAsnAlaSerLysValAlaLsnLysGlyLysSerLys 140  
 Db 361 GCACCTCTGGAGAACCAACGAAABAGCACACCAAAAGTGGCTTAATAAGGABAAAGCAA 420  
 RESULT 12  
 AA243928  
 ID AA243928 standard; cDNA; 1038 BP.  
 XX  
 AC AA243928;  
 XX  
 DT 17-MAR-2000 (first entry)  
 XX  
 DE Murine CID cDNA.  
 XX  
 KM CID; murine; apoptosis; tumour; gene therapy; treatment; se.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FH CDS 76..501  
 FT /\*tag= a  
 FT /product= "CID"  
 XX  
 PN DE19824811-A1.  
 XX  
 PD 09-DEC-1999.  
 XX  
 PF 03-JUN-1998; 98DE-01024811.  
 XX  
 PR 03-JUN-1998; 98DE-01024811.  
 XX  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
 XX  
 PI Rothbarth K, Stammer H, Werner D;  
 XX  
 DR WPI; 2000-063506/06.  
 DR P-PSDB; AAY51025.  
 XX  
 PT Inducing apoptosis by overexpressing the CID gene, particularly for  
 PT treating tumors.  
 XX  
 PS Claim 3; Fig 2; 10pp; German.  
 XX  
 CC This invention describes a novel method for inducing apoptosis which  
 CC comprises overexpressing the CID gene (1). The method is particularly  
 CC used to treat tumors and can also be used in gene therapy. The method has  
 CC no side effects on normal cells (contrast known methods of inducing  
 CC apoptosis such as cytotoxins and radiation), and may be effective on  
 CC cells resistant to conventional treatments. Overexpression of (1) is  
 CC sufficient itself to induce apoptosis but the effect may be increased  
 CC when used in combination with other anti-tumor methods. When cells  
 CC transfected with (1) undergo apoptosis, they release factors that kill  
 CC neighboring, non-transfected cells (bystander effect). This sequence  
 CC encodes the murine CID protein described in the method of the invention  
 XX  
 SQ Sequence 1038 BP; 323 A; 163 C; 219 G; 333 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2.88e-70 Length: 1038  
 Score: 646.00 Matches: 126  
 Percent Similarity: 95.00% Conservative: 7  
 Best Local Similarity: 90.00% Mismatches: 7

Query Match: 90.35% Indels: 0  
DB: 3 Gaps: 0  
US-09-701-618a-2 (1-141) x AAL43975 (1-1038)

QY 1 MetAlaGluGluGluLeuAsnGluAspTyrProValGluLeuHisGluTyrLeuSerAla 20  
DB 76 ATGGGACGGTGAAGAATGATGAATGATTTATCCCGTGAATTCACGAGTCTTTAAACGCC 135  
QY 21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40  
DB 136 CTGGAGAGACTCCCTGGGCTGCTGGACGACATGCTGAAGACCATGATGCGCTGTTCTAGA 195  
QY 41 AsnGluLeuLeuGluLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla 60  
DB 196 AACGAGTTGTGCAAGAGTTGACCCATGTGAACACCAAGGAGTTGATTTGCTGCA 255  
QY 61 TyrThrLeuAsnSerMetPheTyrValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80  
DB 256 TACACCTTAATTAATGATGTTTGGGTTTATTTGGCACTCAAGAGTTAATCCCAAGAG 315  
QY 81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluLe 100  
DB 316 CATCCAGTGAAGCGAAGTGAAGAAAGATCAAGTCTAATGAAACGAGTTAAAGAAATA 375  
QY 101 ThrAspLysLysLysAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn 120  
DB 376 ACAGACAGAAAGAGGCTGCCAAGCTGACAGAGGCTGCTTCAGATTGTCAGAAAG 435  
QY 121 AlaLeuTyrGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140  
DB 436 GCACTCTGGGAACCCAAAGCAAAAGCAACAAAGAGGCTTAATTAAGGAAAGCAAA 495

RESULT 13  
AAL43975  
ID AAL43975 standard; cDNA, 420 BP.

AC AAL43975;  
DT 27-SEP-2002 (first entry)  
XX Mammalian nuclear receptor cofactor CF8 coding sequence.  
DE Mammalian; gene; ss; nuclear receptor cofactor; CF7; CF8;  
XX metabolism regulation; cell homeostasis; cell proliferation;  
KW differentiation; pathological cellular aberration;  
KM cellular defence mechanism.  
XX Mammalia.

OS  
FH Key Location/Qualifiers  
FT CDS 1..420  
FT /\*tag= a  
FT /product= "Mammalian CF8 protein"  
FT /note= "No start or stop codon is given"

W0200242322-A2.

30-MAY-2002.

21-NOV-2001; 2001MO-EP013548.

21-NOV-2000; 2000EP-00125524.

(LION-) LION BIOSCIENCE AG.

PI Jackson D, Casari G, Suckow J;

WPI; 2002-566559/60.

P-PSDB; AAO15406.

Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators

PT useful for inhibiting cellular function of cofactor and for treating  
PT metabolic disorders, immunological indications and hormonal dysfunctions.  
XX Claim 1; Fig 1; 68pp; English.

XX The invention comprises the amino acid and coding sequences of two  
CC mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and  
CC CF8 protein sequences of the invention are useful for screening agents  
CC and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes  
CC involved in cellular functions, such as: regulation of metabolism and  
CC cell homeostasis, cell proliferation and differentiation, pathological  
CC cellular aberrations, or cellular defence mechanisms. The present cDNA  
CC sequence encodes the mammalian nuclear receptor cofactor CF8 protein

Sequence 420 BP; 153 A; 62 C; 100 G; 105 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.5e-70 Length: 420  
Score: 644.00 Matches: 128  
Percent Similarity: 92.86% Conservative: 2  
Best Local Similarity: 91.43% Mismatches: 10  
Query Match: 90.07% Indels: 0  
DB: 6 Gaps: 0

US-09-701-618a-2 (1-141) x AAL43975 (1-420)

QY 2 AlaGlyGluGluLeuAsnGluAspTyrProValGluLeuHisGluTyrLeuSerAlaPhe 21  
DB 1 GCAGCTGAAGAAATTAATGAGACTATCCAGTGAAGAAATTCAGATTATTTGTCAGCAATT 60  
QY 22 GluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArgAsn 41  
DB 61 GCGAATTCATTGAGATGCTGGATGATGATGCTGAAGAAACATGATGCTGTTCTAGAAAT 120  
QY 42 GluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAlaTyr 61  
DB 121 GAGTTGTTGCAGAAATTTGACCCCTTGACACCAAAAGGATTTGGTTCGATAC 180  
QY 62 ThrLeuAsnSerMetPheTyrValTyrLeuAlaThrGlnGlyValAsnProLysGluHis 81  
DB 181 AATATTAAATTCATATGTTTGGCTTATTTGGCACTCAAGAGTGAATCTTAAGAAACAT 240  
QY 82 ProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluLeuHis 101  
DB 241 TCAGTAAAGCAGAAATTCGAAAGATCAAGTATATATGAACAGAGTCAAGAAATAGCA 300  
QY 102 AspLysLysLysAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsnAla 121  
DB 301 GACAGAGAAAGGCTGCCAAGCTGACAGAGGTCGACGCTTCAAGATTGTTAGAAATGCC 360  
QY 122 LeuTyrGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLysSer 141  
DB 361 CTCGAGAACCAAAACCGAAAAATGATCCAAAGTTGCTTAATTAAGGAAAGTAAAGT 420

RESULT 14

AAL43976/c

ID AAL43976 standard; cDNA, 420 BP.

AAL43976;

27-SEP-2002 (first entry)

Mammalian nuclear receptor cofactor CF8 reverse complement sequence.

Mammalian; gene; ss; nuclear receptor cofactor; CF7; CF8;  
KW metabolism regulation; cell homeostasis; cell proliferation;  
KM differentiation; pathological cellular aberration;  
XX cellular defence mechanism.

Mammalia.

W0200242322-A2.



Db	255	TACACATTAAATCAATGTTTGGGTTTATTGGCAACCCAGAGGTTAATCTTAAGAA	314
Qy	81	HisProValIysGInGIuLeuGIuaArgIleaArgValIyMetAsnArgValIysGInIle	100
Db	315	CATCCAGTAAACACGGAATTGGAAAGATCAGATATATATGAAACAGAGTCAAGGAAATA	374
Qy	101	ThrAsp	102
Db	375	ACAGAC	380

Search completed: January 5, 2005, 13:29:44  
Job time : 327.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 4, 2005, 07:11:31 : Search time 3828.49 Seconds  
(without alignments)  
9898.765 Million cell updates/sec

Title: US-09-701-618A-3  
Perfect score: 1040  
Sequence: 1 cgaagccgctgcatgagcgtf.....tcataactctgataaataa 1040

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 1821965908 residues  
Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	958	92.1	2674	3	AK028702 Mus muscu
2	952	91.5	2958	3	AK035169 Mus muscu
3	883.8	85.0	1002	2	BB609670 BB609670
4	805.4	77.4	876	6	CB588136 CB588136
5	805.2	77.4	884	7	CE551758 CE551758
6	768.2	73.9	871	5	BUI51822 BUI51822
7	767.4	73.8	820	5	BUS25697 BUS25697
8	694.6	66.8	733	6	CB317510 CB317510
9	692	66.5	731	6	CB574390 CB574390
10	686	66.0	715	6	CB598466 CB598466
11	678.8	65.3	781	1	AI875855 AI875855
12	678.2	65.2	685	2	BB612968 BB612968
13	661.6	63.6	690	5	BQ746434 BQ746434
14	654.2	62.9	671	6	CD352734 CD352734
15	651.6	62.7	772	6	CB320669 CB320669
16	650.2	62.5	657	2	BB627449 BB627449
17	648.4	62.3	678	5	BM936857 BM936857
18	643.6	61.9	836	4	BI694343 BI694343
19	640.4	61.6	667	7	CK794480 CK794480
20	636.2	61.2	642	5	BQ564508 BQ564508
21	634	61.0	668	6	CD773363 CD773363
22	632.6	60.8	737	4	BI697211 BI697211
23	631.6	60.7	753	6	CB571517 CB571517
24	615	59.1	712	6	CB950526 CB950526

25	612.2	58.9	617	4	BG081538 BG081538
26	579	55.7	640	6	CD766200 CD766200
27	562.6	54.1	648	2	AM611366 AM611366
C 28	561.2	54.0	661	2	AM986301 AM986301
C 29	557.4	53.6	659	6	CD775407 CD775407
30	557.2	53.6	562	4	BM125918 BM125918
31	552.2	53.1	657	1	AV002444 AV002444
32	543.4	52.2	750	6	CB572493 CB572493
33	542	52.1	761	6	CB953171 CB953171
C 34	541.4	52.1	618	6	BE688895 BE688895
35	540.2	51.9	766	6	CB574401 CB574401
36	535.2	51.5	629	6	CD773175 CD773175
37	534.4	51.4	544	2	AM475291 AM475291
38	533.2	51.3	549	1	AA153178 AA153178
39	528.2	50.8	650	6	CP110457 CP110457
40	527.2	50.7	532	4	BM070131 BM070131
41	527.2	50.7	781	4	BG862190 BG862190
42	525.2	50.5	534	1	AI786504 AI786504
43	521.2	50.1	526	1	AA476164 AA476164
44	519.2	49.9	795	8	AF046558 AF046558
45	519.2	49.9	795	9	CG478915 CG478915

#### ALIGNMENTS

RESULT 1  
AK028702  
LOCUS  
DEFINITION  
Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:4732439J08 product:nuclear DNA binding protein, full insert sequence.

ACCESSION  
AK028702  
VERSION  
AK028702.1 GI:26324641  
KEYWORDS  
HTC, CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus

REFERENCE  
AUTHORS  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.  
TITLE  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL  
MEDLINE  
99279253  
PUBMED  
10349636

REFERENCE  
AUTHORS  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.  
TITLE  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL  
MEDLINE  
20499374  
PUBMED  
11042159

REFERENCE  
AUTHORS  
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsumi,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakauechi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,Y., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsubara,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.  
TITLE  
RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
JOURNAL  
MEDLINE  
20530913  
PUBMED  
11076861

REFERENCE  
AUTHORS  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
TITLE  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
JOURNAL  
MEDLINE  
11076861  
PUBMED  
11076861  
AUTHORS  
The FANTOM Consortium and the RIKEN Genome Exploration Research

TITLE Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
REFERENCE 6 (bases 1 to 2674)  
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Osada, N.,  
Ozaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takeku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M., and Hayashizaki, Y.  
TITLE Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
COMMENT A cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
location/Qualifiers  
source  
1. 2674  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6J"  
/db\_xref="PANTOM:DB:4732439J08"  
/db\_xref="taxon:10090"  
/clone="4732439J08"  
/tissue\_type="skin"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days neonate"  
58. 483  
/note="unnamed protein product; nuclear DNA binding  
protein (MGI:MGI:1927354, GB|NM\_020558, evidence: BLASTN,  
98%, match=1021)  
putative"  
/codon\_start=1  
/protein\_id="BAC26075.1"  
/db\_xref="GI:26324642"  
/translation="MAGEBNEDYPIVEIHSLTALSSLGAVDMKTMVAASRNELT  
OKDPRLOAKVDVSAFTLNSMPVYLAQGVNPKHEHPKOELEIRIYVNNVKEITD  
KKQAALDRGAARFPVKMLMEPKASTPKVANKSKK"  
2661. 2666  
/note="putative"  
2674  
/note="putative"  
ORIGIN  
polya\_site  
Query Match 92.1%; Score 958; DB 3; Length 2674;  
Best Local Similarity 98.6%; Pred. No. 1.6e-205;  
Matches 1008; Conservative 0; Mismatches 10; Indels 4; Gaps 4;  
Db  
QY 22 ATCATGTCGACCTATTTCCCGAGAGACGGCTCCACGGTTTGAAGTTGTCACAAATG 81  
|||  
Db 2 ATCATGTCGACCTATTTCCCGAGAGACGGCTCCACGGTTTGAAGTTGTCACAAATG 61  
QY 82 CAGGTAAAGAAATGAATGAATATTCCTGAGAAATTCACAGATCTTAAACAGCCCTGG 141  
|||  
Db 62 CAGGTAAAGAAATGAATGAATATTCCTGAGAAATTCACAGATCTTAAACAGCCCTGG 121

QY 142 AGAGCTCCCTGGGTCGTGTGGAGACATGCTGAAGACCATGATGCTGTTCTGAGAAAGC 201  
Db 122 AGAGCTCCCTGGGTCGTGTGGAGACATGCTGAAGACCATGATGCTGTTCTGAGAAAGC 181  
QY 202 AGTGTTCACAAAGTGGAGCCATTTGAGACAGAAAGGTGGATTTAGTTTCTGCATACA 261  
Db 182 AGTGTTCACAAAGTGGAGCCATTTGAGACAGAAAGGTGGATTTAGTTTCTGCATACA 241  
QY 262 CCTTAATTCATGTTTGGGTTTATTTGGCACTCAAGAGGTTAAATCCCAAGACATC 321  
Db 242 CCTTAATTCATGTTTGGGTTTATTTGGCACTCAAGAGGTTAAATCCCAAGACATC 301  
QY 322 CAGTGAAGCAGAACTGGAAGAAATGAGTCTACATGAGACAGATTAAAGAAATACAG 381  
Db 302 CAGTGAAGCAGAACTGGAAGAAATGAGTCTACATGAGACAGATTAAAGAAATACAG 361  
QY 382 ACAAGAAAGAGGCTGCCAAGCTGGAAGAGTCTCTCGAGATTGTCAAGAAAGGCAC 441  
Db 362 ACAAGAAAGAGGCTGCCAAGCTGGAAGAGTCTCTCGAGATTGTCAAGAAAGGCAC 421  
QY 442 TCTGGGAACCCAGCAAGCAAAAGCAACCAAAAGTGTCTAATAAGGAAAGCAAACT 501  
Db 422 TCTGGGAACCCAGCAAGCAAAAGCAACCAAAAGTGTCTAATAAGGAAAGCAAACT 481  
QY 502 AATCTTTGGTTTGAATGATGATGATGTTTCAAAAGATCATCTTTTATCATGTTTACA 561  
Db 482 AATCTTTGGTTTGAATGATGATGATGTTTCAAAAGATCATCTTTTATCATGTTTACA 541  
QY 562 TGTAGTTAGTACCAATGAGTGTGTTAAATGATGATCTTTTGAATCATGATTAATTT 621  
Db 542 TGTAGTTAGTACCAATGAGTGTGTTAAATGATGATCTTTTGAATCATGATTAATTT 601  
QY 622 ACACATTCATTTGATGATGATGATCTTTTGTGCTGAGAAAGTTAAGTGTCTTGT 681  
Db 602 ACACATTCATTTGATGATGATGATCTTTTGTGCTGAGAAAGTTAAGTGTCTTGT 661  
QY 682 TGAATTTCAATTAAGCATCATGATGTTTAAATGTTAAGATATTCATTAAGCAGTTG 741  
Db 662 TGAATTTCAATTAAGCATCATGATGTTTAAATGTTAAGATATTCATTAAGCAGTTG 721  
QY 742 TGAATTTCAATTAAGCATCATGATGTTTAAATGTTAAGATATTCATTAAGCAGTTG 801  
Db 722 TGAATTTCAATTAAGCATCATGATGTTTAAATGTTAAGATATTCATTAAGCAGTTG 781  
QY 802 AGTGTCTATCTGTAACCTGAGCTGTAAGGACATTTGTTTCAGTAATGATGAGAAAT 861  
Db 782 AGTGTCTATCTGTAACCTGAGCTGTAAGGACATTTGTTTCAGTAATGATGAGAAAT 840  
QY 862 ACAGTGAATTAATACCACTGCTGTTCTGTTCAAGTATTTCAAGTGTGTTGTTGATTT 920  
Db 841 ACAGTGAATTAATACCACTGCTGTTCTGTTCAAGTATTTCAAGTGTGTTGTTGATTT 900  
QY 921 TTTTGTGTGTGTGAGTAATCTGTCTGTAATTTCAAGTATTTCAAGTGTGTTGATTT 980  
Db 901 TTTTGTGTGTGTGAGTAATCTGTCTGTAATTTCAAGTATTTCAAGTGTGTTGATTT 960  
QY 979 CTGACTTTAATCTTCATGTTCCATTAATAAATTAATGTTCTCATTAATCTGATGAA 1038  
Db 961 CTGACTTTAATCTTCATGTTCCATTAATAAATTAATGTTCTCATTAATCTGATGAA 1020  
QY 1039 AA 1040  
Db 1021 AA 1022  
RESULT 2  
AK035169  
LOCUS  
DEFINITION  
AK035169 2958 bp mRNA linear HTC 03-APR-2004  
and neck cDNA, 12 days embryo embryonic body between diaphragm region  
product:nuclear DNA binding protein, full insert sequence.  
ACCESSION  
AK035169  
AK035169.1 GI:26330473

KEYWORDS	HTG, CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE	1
AUTHORS	Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Ktsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsubara, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
MEDLINE	11076861
PUBMED	11076861
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	12000000
PUBMED	12000000
REFERENCE	6 (bases 1 to 2958)
AUTHORS	Adachi, S., Aizawa, K., Akiyama, T., Aizawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagata, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, P., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanakawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koye, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohnishi, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akiba, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Yuramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.jp, url: http://genome.resgsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. url: http://genome.resgsc.riken.jp/ url: http://fantom.resgsc.riken.jp/ Location/Qualifiers
FEATURES	

Source	1..2958
	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="C57BL/6J"
	/db_xref="FANTOM DB:9430096A21"
	/db_xref="Ensembl:ENST00000263967.5"
	/clone="9430096A21"
	/tissue_type="embryonic body between diaphragm region and neck"
	/clone_1db="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="12 days embryo"
	52..477
	/note="unmated protein product; nuclear DNA binding protein (MCD MGI:1927354, GB NM_020558, evidence: BLASTN, 98%, match=1021)
	putative"
	/codon_start=1
	/protein_id="BAC28967.1"
	/db_xref="GI:26330474"
	/translation="MAGEBMDYDVEIHSLRLBSSLGAVDDMLKTPMAVSRNELLQKIDPLEQAVDVLSAVTLNMFVYATGVKPEKIPVQOEIRLRVYNNRVEITD KKKAAKIDRGASRFVFNALMBEPKAKSTPRVANKGSKH"
ORIGIN	
Query Match	91.5%; Score 952; DB 3; Length 2958;
Best Local Similarity	98.6%; Pred. No.3.7e-204;
Matches 1002; Conservative	0; Mismatches 10; Indels 4; Gaps 4;
Db	28 GTGGACCTATATTCGGGAGACAGACGCGTATTTGATGGTCAACATGACAGGTG 87
	2 GTGGACCTATATTCGGGAGACAGACGCGTATTTGATGGTCAACATGACAGGTG 61
QY	88 AAGAAATGAATGAAGATTTATCCCTAGAAATTCACGAGTCTTTAACAGCCCTGAGAGCT 147
	62 AAGAAATGAATGAAGATTTATCCCTAGAAATTCACGAGTCTTTAACAGCCCTGAGAGCT 121
QY	148 CCTGGGTGCTGTGGACGACATGCTGAAGACATGATGCTGTTCTTGAAGACAGTGTG 207
	122 CCTGGGTGCTGTGGACGACATGCTGAAGACATGATGCTGTTCTTGAAGACAGTGTG 181
QY	208 TGCAGAGTTGGACCCATTTGGAAACAACAAAGGTGATTTAGTTCTGCATACACCTTAA 267
	182 TGCAGAGTTGGACCCATTTGGAAACAACAAAGGTGATTTAGTTCTGCATACACCTTAA 241
QY	268 ATTCAATGTTTGGGTTTATTTGGCACTCAAGAGTTAATCCCAAGACATCACTGA 327
	242 ATTCAATGTTTGGGTTTATTTGGCACTCAAGAGTTAATCCCAAGACATCACTGA 301
QY	328 AGCAGAACTGGAAGAATCAGACTTACATGAACAGAGTTAAAGAAATTAACAGACAGA 387
	302 AGCAGAACTGGAAGAATCAGACTTACATGAACAGAGTTAAAGAAATTAACAGACAGA 361
QY	388 AGAAGGTGCGCAAGCTGACAGAGTCTGCTCGAATTTGTCANAGAAGCACTCGGG 447
	362 AGAAGGTGCGCAAGCTGACAGAGTCTGCTCGAATTTGTCANAGAAGCACTCGGG 421
QY	448 AACCCAAACGAAAGACACAAAGTGGCTAATAAAGGAAAGCAAACTAATCTT 507
	422 AACCCAAACGAAAGACACAAAGTGGCTAATAAAGGAAAGCAAACTAATCTT 481
QY	508 TTGGTTTGTATGATCATGTTTCAAAAAGTACATCTTTTAAATCAAGTTTAAATGTA 567
	482 TTGGTTTGTATGATCATGTTTCAAAAAGTACATCTTTTAAATCAAGTTTAAATGTA 541
QY	568 TATGTGACATGTGTGTGTTAAATGATTTCTTTTGAATTCATGTATAATTTACAT 627
	542 TATGTGACATGTGTGTGTTAAATGATTTCTTTTGAATTCATGTATAATTTACAT 601
QY	628 TACATTTGTATGATCATGTTTCAAAAAGTACATCTTTTAAATCAAGTTTAAATGTA 687
	602 TACATTTGTATGATCATGTTTCAAAAAGTACATCTTTTAAATCAAGTTTAAATGTA 661
QY	688 TCATATAAGCATCATGATGTTTAAATATTTGTAAGATTTCTATAAGCAGTTGTGAAT 747

```
Db      662 TCATTAAGACATCATGATGTTTAAATTTGTAAGATTTCTATAAGCAGTTGTGAAT 721
Qy      748 CCAATATGTTCTCTGTAAACATTGTAGTGTGTAATGAACAATGATATATGAAGTGTG 807
Db      722 CCAATATGTTCTCTGTAAACATTGTAGTGTGTAATGAACAATGATATATGAAGTGTG 781
Qy      808 CTATCTGTAGACCTCGAGGTGTAGAGCATTTGTTTCAATATGATGAGAAATACAGTG 867
Db      782 CTATCTGTAGAC-CTGAGGTGTAGAGCATTTGTTTCAATATGATGAGAAATACAGTG 840
Qy      868 ACTTAATATACCACTGTG-TTCTGTTCAGTTAGTTCACATGTTTCGTGATTTTTTT 926
Db      841 ACTTAATATACCACTGTGTTTCTGTTCAGTTAGTTCACATGTTTCGTGATTTTTTT 900
Qy      927 TTTTGTAGTAAAT-CTGTCTGATATTCAAAT-CAAAATGAAACCTTAAGCTGTAC 984
Db      901 TTTTGTAGTAAATGTTGTTGTTGATATCAAACTCAAAATGAAACCTTAAGCTGTAC 960
Qy      985 TTTAATTTCTTCAATGTTTCAATTTAAATAAATGTTCTCATTAAGTGTGAAAA 1040
Db      961 TTTAATTTCTTCAATGTTTCAATTTAAATAAATGTTCTCATTAAGTGTGAAAA 1016
```

```
RESULT 3
BB609670 1002 bp mRNA linear EST 26-OCT-2001
LOCUS BB609670 RIKEN full-length enriched, 18 days embryo Mus musculus
DEFINITION cDNA clone 1110036E10 5', mRNA sequence.
ACCESSION BB609670
VERSION BB609670.1 GI:16451385
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
REFERENCE 1 (bases 1 to 1002)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
Kono,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,P.,
Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Shuho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
```

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

Wajiki,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matsubara,S., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamakata,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and

#### FEATURES

source

Hayashizaki,Y.  
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
e mouse tissues.

#### location/Qualifiers

```
1..1002
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="1110036E10"
/dev_stage="18 days embryo"
/lab_host="SOLR"
/clone_lib="RIKEN full-length enriched, 18 days embryo"
/note="Site_1: XhoI; Site_2: SctI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGGAGAGAGATCCCAAGAGCTCATTAATTAATTAACCCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGGAGAGAGATCCCAAGAGCTCATTAATTAATTAACCCCCCCCC 3']. cDNA was cleaved with XhoI and SctI."
```

#### ORIGIN

Query Match 85.0%; Score 883.8; DB 2; Length 1002;  
Best Local Similarity 97.6%; Pred. No. 8,2e-189;  
Matches 929; Conservative 0; Mismatches 18; Indels 5; Gaps 3;

```
Qy      1 CAGAAGCCGTGCATGAGCGCTCATCGTGGACCTATTTCCGAGACAGCGCTCCAG 60
Db      29 CAGAAGCCGTGCATGAGCGCTCATCGTGGACCTATTTCCGAGACAGCGCTCCAG 88
Qy      61 GTATTGAGTTGTCACAAATGCGAGTGAAGAAATGAATGAATATTCGGTAGAAATTC 120
Db      89 GTATTGAGTTGTCACAAATGCGAGTGAAGAAATGAATGAATATTCGGTAGAAATTC 148
Qy      121 ACGAGCTTTTAAACGCCCTGAGAGAGCTCCCTGGGTGCTGTGAGACGACATGTAAGCA 180
Db      149 ACGAGCTTTTAAACGCCCTGAGAGAGCTCCCTGGGTGCTGTGAGACGACATGTAAGCA 208
Qy      181 TGATGGCTGTTTCTAGAAACGAGTTGTGCAAGATTGACCCCATGGAAACAAGCAAG 240
Db      209 TGATGGCTGTTTCTAGAAACGAGTTGTGCAAGATTGACCCCATGGAAACAAGCAAG 268
Qy      241 TGGATTTAGTTTTCGATACACCTTAATTCATGTTTGGTTTATTTGGCACTCAAG 300
Db      269 TGGATTTAGTTTTCGATACACCTTAATTCATGTTTGGTTTATTTGGCACTCAAG 328
Qy      301 GAGTTAATCCCAAGACATTCAGTGAAGCAAGTGAAGAAATCAAGTCTTAATGTA 360
Db      329 GAGTTAATCCCAAGACATTCAGTGAAGCAAGTGAAGAAATCAAGTCTTAATGTA 388
Qy      361 ACAAGTTTAAAGAAATTAACGACGACGAGAGAGGCTCCCAAGCTGACAGAGTGTCTT 420
Db      389 ACAAGTTTAAAGAAATTAACGACGACGAGAGAGGCTCCCAAGCTGACAGAGTGTCTT 448
Qy      421 CGAGATTGTCAAGAAAGGCACTCGGGAACCCAAAGAAAGCAACCAAAAGTGCTTA 480
Db      449 CGAGATTGTCAAGAAAGGCACTCGGGAACCCAAAGAAAGCAACCAAAAGTGCTTA 508
Qy      481 ATTAAGGAAAGCAACCTAATCTTTGTGTTTATGATGATGATTTCAAAAAGTACA 540
Db      509 ATTAAGGAAAGCAACCTAATCTTTGTGTTTATGATGATGATTTCAAAAAGTACA 568
Qy      541 TCCCTTTTATGATGATTTAATGATGATTTGACCATGTGTGTTAATGATTCCTT 600
```

Db	569	TCCTTTTAACAGTTTCAAGATGTAAGTATGACCACTGCGTGTTTAAATGATTCCTT	628
Qy	601	TTGGATTCATGATTAATAATTACACATTACATTTGTGATCTGAATCTTTTTTGGCTGA	660
Db	629	TTGGAAATTCATGTAATAAATTTTACATTACATTTGTGATTAAGTAATCTTTTTTGGCTGA	688
Qy	661	GAAAGATTAAAGTTGCTCTTTGTGATTTTTCATATAAAGCATATGATGCTTTAATATGT	720
Db	689	GAAAGATTAAAGTTGCTCTTTGTGATTTTTCATATAAAGCATATGATGCTTTAATAATGT	748
Qy	721	AAGATATTTCTATTAAGCAGTTGGAATCAAAATGTTCTCTGTAACATTTGATAGTTTG	780
Db	749	AAGATATTTCTATTAAGCAGTTGGAATCAAAATGTTCTCTGTAACATTTGATAGTTTG	808
Qy	781	AAATGAACAATGATATTTATGAAGTGTGCTATCTGTAGACCTTGAGGTGTAAGACATTTG	840
Db	809	AGATGAACAATGATATTTATGAAGTGTGCTATCTGTAGACCTTGAGGTGTAAGACATTTG	867
Qy	841	TTTTTCAGTAATGATGAGAAATACAGTGACTTAATATCCACTCTGTTCTGTTCAGTTTG	900
Db	868	TTTTTCAGTAATGATGAGAAATACAGTGACTTAATATCCACTCTCTTCTGTTCAGTTTG	927
Qy	901	TTCAACAATGTTGCGATTTTTTTTTTTTTTTTGAATAATCTGCTGCTGAAT	952
Db	928	TTCAACAATGTTGCGATTTTTTTTTTTTTTTTGAATAATGCTGCTGGAAT	975

RESULT 4					
CB588136					
LOCUS	CB588136	876 bp	mRNA	linear	EST 03-APR-2007
DEFINITION	AGNCNCURT 13003966 NIH MGC 136 Mus musculus cDNA clone				
	IMAGE:30253578 5', mRNA sequence.				

ORGANISM	Mus musculus
SOURCE	Mus musculus (house mouse)
KEYWORDS	EST.
VERSION	CB588136.1
ACCESSION	GI:29505992
CB588136	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Mus.	1 (bases 1 to 876)	NIH-MGC <a href="http://mgc.nhl.nih.gov/">http://mgc.nhl.nih.gov/</a> .	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)
Contact: Robert Stransberg, Ph.D.				

Email: [csapbs-remail.nih.gov](mailto:csapbs-remail.nih.gov)  
Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MSC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ULNL at:  
<http://image.lnl.gov>  
Plate: NDAM339 row: e column: 11  
High quality sequence start: 7  
High quality sequence stop: 704.

FEATURES	Location/Qualifiers
source	1. .876

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone IMAGE:30293578"
/tissue type="embryonic limb, maxilla and mandible"
/lab host="DH10B (phage-resistant)"
/clone.lib="NH_MGC_136"
/notes="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTAGTCTGATCGACGCGCCGCT(7)15-3',
size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by Reegen, Invitrogen

```

ORIGIN	Corp. Note: this is a NIH_MGC Library."
Query Match	77.4%; Score 805.4; DB 6; Length 876;
Best Local Similarity	96.8%; Pred. No. 4.1e-171;
Matches 843; Conservative	0; Mismatches 26; Indels 2; Gaps 2;

OY	16	GGCGTCATCATCGTGCACCTATTTC	CCCGAGACAGGCGCTCCACGGTATTGAGTTGGTCA	75	
Db	6	GTCCGTCAATCATCGTGA	CCCTATTTC	CCCGAGACAGGCGCTCCACGGTATTGAGTTGGTCA	65
OY	76	CAATGCGACGGTGAAAGAAATGA	-ATGAAGATTATTC	CCCGTAGAAATTCACGAGTCTTTTACA	134
Db	66	CAATGCGACGGTGAAAGAAATGA	TGTGAAGATTATTC	CCCGTAGAAATTCATGAGTCTTTTACA	125
OY	135	GCCTCTGGAGAGCTCC	CTGGGTGCTGTGGAGACATGCTGGAAGACCAATGATGCGTGTCT	194	
Db	126	GCCTCTGGAGAGCTCC	CTGGGTGCTGTGGAGACATGTTGAAGACCAATGATGCGTGTCT	185	
OY	195	AGAAACGAGTTGTTGC	AGAAAGTTGACCCATTGGACCAACGAAAGGT-GGATTTAGTTTC	253	
Db	186	AGAAACGAGTTGTTGC	AGAAAGTTGACCCATTGGACCAACGAAAGGTGGATTTAGTTTC	245	
OY	254	TGCATACACCTTAATTC	ATGTTTTGGGTTTATTGGGCAATCCAGGATTTAATCCCA	313	
Db	246	TGCATACACCTTAATTC	ATGTTTTGGGTTTATTGGGCAATCCAGGATTTAATCCCA	305	
OY	314	AGACATCCAGTGAAGGAGGAACTG	GAAGAAATCAGAGTCTACATGAAACAGAGTTTAAAGA	373	
Db	306	AGACATCCAGTGAAGGAGGAACTG	GAAGAAATCAGAGTCTACATGAAACAGAGTTTAAAGA	365	
OY	374	AATTAACAGACAAAGAAAGGCTG	CAAGCTGGAACAGAGTTGTCAA	433	
Db	366	AATTAACAGACAAAGAAAGGCTG	CAAGCTGGAACAGAGTTGTCAA	425	
OY	434	GAAGGCACTCTGGGAAACCC	CAAGAAAGACACCCAAAGTGCTATTAAGGGAAG	493	
Db	426	GAAGGCACTCTGGGAAACCC	CAAGAAAGACACCCAAAGTGCTATTAAGGGAAG	485	
OY	494	CAAAACATTAATCTTTGGTTTG	TATGATACATGTTTTCAAAAGATACATCCTTTTAAATCA	553	
Db	486	CAAAACATTAATCTTTGGTTTG	TATGATACATGTTTTCAAAAGATACATCCTTTTAAATCA	545	
OY	554	GTTTACATATGATGATATG	TACCATGATGATGTTTAAATGATTTCCTTTGGAAATTCATGT	613	
Db	546	GTTTACATATGATGATATG	TACCATGATGATGTTTAAATGATTTCCTTTGGAAATTCATGT	605	
OY	614	ATAAATTTAACATTA	CATTGTGATATCTGAATCTTTTTTCTGAGAAAGATTAAAGTT	673	
Db	606	TTAAATTTAACATTA	CATTGTGATATCTGAATCTTTTTTCTGAGAAAGATTAAAGTT	665	
OY	674	GTCCTTTGTGATTTTCA	TATTAAGACATCAGATGTTTAAATTTGTAAGATATTTCTATA	733	
Db	666	GTCCTTTGTGATTTTCA	TATTAAGACATCAGATGTTTAAATTTGTAAGATATTTCTATA	725	
OY	734	AGCAGTTGTAATCCAAATG	TCTCTGTAAACATTTGTGTTTGAATGAACAATGA	793	
Db	726	AGCAGTTGTAATCCAAATG	TCTCTGTAAACATTTGTGTTTGAATGAACAATGA	785	
OY	794	TATTTAGAAAGTGCTAT	CTGTGAGACCTCGAGGTGTAAAGACATTTGTTTCAGTAATGA	853	
Db	786	TATTTAGAAAGTGCTAT	CTGTGAGAAAGGTGTAAAGACATTTGTTTCAGTAATGA	845	
OY	854	TGAGAAATTAACAGTAA	ATATACCCACTT	884	
Db	846	TGAGAAATTAACAGTAA	ATATACCCACTT	876	

RESULT 5	
CE551758	
LOCUS	CE551758
DEFINITION	CE551758 864 bp mRNA linear EST 25-SEP-2003
	AGAGCOURT_15587320 NIH_MGC_222 Mus musculus cDNA clone
	IMAGE:30521123 5', mRNA sequence.

ACCESSION CF551758  
 VERSION CF551758.1 GI:34888592  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 864)  
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Daniela S. Gerhard, Ph.D.  
 COMMENT Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: Naryan Bhat  
 cDNA Library Preparation: Express Genomics  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
 Clone distribution: MGC clone distribution information  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: NDAM595 row: f column: 12  
 High quality sequence stop: 769.  
 Location/Qualifiers  
 1..864

FEATURES  
 source  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30521123"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH-MGC-222"  
 /note="Organ: Placenta; Vector: pExpress-1; Site 1: EcoRV, Site 2: NotI; RNA obtained from three placentas from female C57/Bl6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer.  
 5'-pgactgagctctgagatcgccgagccgccc(7)25-3' and cloned into the EcoRV/NotI sites of pExpress-1. Size-selection >1 kb not amplified. (Normalized version of this library is NIH-MGC-203.) Library constructed by Express Genomics (Frederick, MD)."

## ORIGIN

Query Match 77.4%; Score 805.2; DB 7; Length 864;  
 Best Local Similarity 97.2%; Pred. No. 4.5e-171;  
 Matches 830; Conservative 0; Mismatches 23; Indels 1; Gaps 1;

51 GGGCCCGCGGATTGAGTGGTCAATGCGAGTGAAGAATGAATGATATCC 110  
 1 GGGCCCGCGGATTGAGTGGTCAATGCGAGTGAAGAATGAATGATATCC 60  
 111 GTAGAAATTCAGAGCTTTTAACGCGCTGGAAGCTCCCTGGGTCTGTGACGATG 170  
 61 GTAGAAATTCAGAGCTTTTAACGCGCTGGAAGCTCCCTGGGTCTGTGACGATG 120  
 171 CTGAAGCCATGATGGCTGTTCTAGAAAGAGTGTGGAAGAGTTGAGCCATTGAA 230  
 121 CTGAAGCCATGATGGCTGTTCTAGAAAGAGTGTGGAAGAGTTGAGCCATTGAA 180  
 231 CAAGCAAGGTGATTTAGTTTCTGATACACCTTAAATCAATGTTTGGTTATTTG 290  
 181 CAAGCAAGGTGATTTAGTTTCTGATACACCTTAAATCAATGTTTGGTTATTTG 240  
 291 GCAACTCAAGGATTAATCCCAAGAGCATCCAGTGAAGCAAGAAAGTGAAGATCGA 350  
 241 GCAACTCAAGGATTAATCCCAAGAGCATCCAGTGAAGCAAGAAAGTGAAGATCGA 300  
 351 GTCTACATGAACGAGTTAAAGAAATACAGCAAGAAAGGCTCCAGACTGACAGCA 410  
 |||||

Db 301 GTCTACATGAACGAGTTAAAGAAATACAGCAAGAAAGGCTCCAGACTGACAGCA 360  
 QY 411 GGTGCTGCTTGAGATTGTCAAGAGGACCTCTGGGAACCCAAAGCAAGCAACCA 470  
 Db 361 GGTGCTGCTTGAGATTGTCAAGAGGACCTCTGGGAACCCAAAGCAAGCAACCA 420  
 QY 471 AAAGTGGCTAATAAAGGAAAGCAACACTAATCTTTGGTTTGAATGATGTTTC 530  
 Db 421 AAAGTGGCTAATAAAGGAAAGCAACACTAATCTTTGGTTTGAATGATGTTTC 480  
 QY 531 AAAAGTACATCTTTTAAATCACTTACATGATGATGATGATGATGATGATGATG 590  
 Db 481 AAAAGTACATCTTTTAAATCACTTACATGATGATGATGATGATGATGATGATG 540  
 QY 591 TGAATTCCTTTTGAATTCATGATGATGATGATGATGATGATGATGATGATGATG 650  
 Db 541 TGAATTCCTTTTGAATTCATGATGATGATGATGATGATGATGATGATGATGATG 600  
 QY 651 TTTTGGCTGAGAAAGATTAGTGTCTTTGTTGATTTCAATTAAGATCATGATGAT 710  
 Db 601 TTTTGGCTGAGAAAGATTAGTGTCTTTGTTGATTTCAATTAAGATCATGATGAT 660  
 QY 711 TTAATATGTAAGATATCTTAAGAGAGTGTGAATCCAAATGTTCTCTGTAACAT 770  
 Db 661 TTAATATGTAAGATATCTTAAGAGAGTGTGAATCCAAATGTTCTCTGTAACAT 720  
 QY 771 GTAGTGTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 830  
 Db 721 GTAGTGTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 779  
 QY 831 AGACACTTTGTTTTCAGTAATGATGAGAAATACAGTGAATTAATACCACTGTTCT 890  
 Db 780 AGACACTTTGTTTTCAGTAATGATGAGAAATACAGTGAATTAATACCACTGTTCT 850  
 QY 891 GTTCACTTGTGCA 904  
 Db 840 TTCTGTTCAGTTTA 853

RESULT 6  
 BU51822 871 bp mRNA linear EST 03-SEP-2002  
 LOCUS AGENCOURT 8754213 NIH-MGC\_130 Mus musculus cDNA clone IMAGE:6332444  
 DEFINITION 5', mRNA sequence.  
 BU51822  
 ACCESSION BU51822.1 GI:22665354  
 VERSION  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 871)  
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgaabs-remail.nih.gov  
 Tissue Procurement: Mark Macconochie, Ph.D. and Nancy L. Freeman, Ph.D.  
 cDNA Library Preparation: ResGen, Invitrogen Corp  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
 Clone distribution: MGC clone distribution information  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM13790 row: a column: 21  
 High quality sequence stop: 615.  
 Location/Qualifiers  
 1..871

## FEATURES

source  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6332444"



ORIGIN

[illegible]

ACCESSION	IMAGE:6534234 5', mRNA sequence.
VERSION	BUS25697
KEYWORDS	BUS25697.1 GI:22836150
SOURCE	EST.
ORGANISM	Mus musculus (house mouse)
REFERENCE	Mus musculus
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus
JOURNAL	1 (baaes 1 to 820)
COMMENT	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:straus@ncl.nih.gov">straus@ncl.nih.gov</a>

Tissue Procurement: Dr. Michael Brownstein  
 cDNA Library Preparation: Michael Brownstein Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCM2692 row: a column: 18  
 High quality sequence stop: 581.  
 Location/Qualifiers

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/_clone="IMAGE:6534234"
/_lab_host="DH10B (T1-phage-resistant)"
/_clone_lib="NIH_MGC_14"
/_note="Organ: Brain; Vector: pDNR-VIB; Site.1: SfiI
(gggcatatggcc) / Site.2: SfiI (ggccggccggcc) / cDNA made
by oligo-dt priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AACCATGCTATTCACGACAGAGGCGCATTCAGCGCCGG-3' and
5'-ATTTCAGAGCGCCGAGCGGCCGACATG-CT(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.2-0.6
kb size fraction (other fractions present in NIH MGC 1433n
library created in the laboratory of W. Brownstein (NIH).
NIH). Note: this is a NIH_MGC Library."

```

Query Match	Similarity	73.8%	Score 767.4	DB 5	Length 820
Best local	Similarity	97.6%	Pred. No. 1,56-162		
Matches 799	Conservative	0	Mismatches 18	Indels 2	Gaps 2
QY	35	CTATTTC	CCGGAGACAGGCGCTCCACGGTATTGAGTTGGTGCACATGCGAGGTGAAGAAAT	94	
Db	4	CTATTTC	CCGGAGACAGGCGCTCCACGGTATTGAGTTGGTGCACATGCGAGGTGAAGAAAT	63	
QY	95	GAATGAAGATTATTC	CCCGTAGAAATTCACGAGTCTTTTAAACAGCCCTGGAGACTCCCTGGG	154	
Db	64	GAATGAAGATTATTC	CCCGTAGAAATTCACGAGTCTTTTAAACAGCCCTGGAGACTCCCTGGG	123	
QY	155	TGCTGTGACGACATGCTGAAGACATGATGGCTGTTCTTGAAACGAGTGTGGCAGAA	214		
Db	124	TGCTGTGACGACATGCTGAAGACATGATGGCTGTTCTTGAAACGAGTGTGGCAGAA	183		
QY	215	GTTGAGCCCATTTGGAAACAAAGGTGAATTTAGTTTGTGCATACCTTTAAATTCAAT	274		
Db	184	GTTGAGCCCATTTGGAAACAAAGGTGAATTTAGTTTGTGCATACCTTTAAATTCAAT	243		
QY	275	GTTTTGGGTTTATTTTGGCACTCAAGAGTTTAAATCCCAAGAGCATCCAGTGAAGCAGGA	334		
Db	244	GTTTTGGGTTTATTTTGGCACTCAAGAGTTTAAATCCCAAGAGCATCCAGTGAAGCAGGA	303		
QY	335	ACTGAAAGATCAGAGTCTACATGAAACAGAGTTTAAAGAAATAACAGACAAAGAAAGGC	394		
Db	304	ACTGAAAGATCAGAGTCTACATGAAACAGAGTTTAAAGAAATAACAGACAAAGAAAGGC	363		
QY	395	TGCCAAGCTGACAGAGGTCGCTTCAGATTGTCAAGAGGCACTCTGGAAACCCA	454		

Db 364 TCCCAAGCTGACGAGAGTCTCTCTCTGAGATTTCAGAGAACGACCTCTGGAAACCAA 423  
Oy 455 ACCAAAAAGCAACCAAAAGTGGCTTAATAAAGGAAAAAGCAACACTATCTTTGGTTT 514  
Db 424 AGCAAAAAGCAACCAAAAGTGGCTTAATAAAGGAAAAAGCAACACTATCTTTGGTTT 483  
Oy 515 TGATGTACATGTCTTCAAAAAGTACATCTTTTATATCATGTTTACATGTATGTATGTA 574  
Db 484 TGATGTACATGTCTTCAAAAAGTACATCTTTTATATCATGTTTCAATGTATGTATGTA 543  
Oy 575 CCATGTGTGTCTTAAATGATTCCTTTTGAATTCATGTATATATTTTACATTAACATTT 634  
Db 544 CCATGTGTGTCTTAAATGATTCCTTTTGAATTCATGTATATATTTTACATTAACATTT 603  
Oy 635 GTGATACATGATCTTTTCTTCTGAGAAAGATTAAGTTGTCTTTGTGATTTTCAATA 694  
Db 604 GTGATACATGATCTTTTCTTCTGAGAAAGATTAAGTTGTCTTTGTGATTTTCAATA 662  
Oy 695 AAGCATCATGATGTCTTAAATGTATAGATATCTTAAAGAGTTGTGAAATCCAAATG 754  
Db 663 AAGCATCATGATGTCTTAAATGTATGTATAGATATCTTAAAGAGTTGTGAAATCCAAATG 722  
Oy 755 TTCTCTGAAACATTTGAGTGTGAAATGAAACATGATATTTATGAAGTGTCTATCTG 814  
Db 723 TTCTCTGAAACATTTGAGTGTGAAATGAAACATGATATTTATGAAGTGTCTATCTG 782  
Oy 815 TAGACCTGAGGTGTAAGACATTTGTTTCAATATGA 853  
Db 783 TAGACCTGAGGTGTAAGACATTTGTTTCAATATGA 820

RESULT 8  
LOCUS CB317510  
DEFINITION ABEINCOURT\_12397793 NIH MGC\_165 Mus musculus cDNA clone  
IMAGE:30284713 5', mRNA sequence.  
ACCESSION CB317510  
VERSION CB317510.1 GI:28841745  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 733)  
NIH-MGC http://mgi.mcg.mcg.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. Leslie L. Heckert  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
http://image.llnl.gov  
Plate: NDCM49 row: d column: 02  
High quality sequence stop: 584.  
Location/Qualifiers  
1. 733  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30284713"  
/tissue\_type="primary cultures of Sertoli cells"  
/lab\_host="DH10B (T1-phage-resistant)"  
/clone\_id="NIH MGC\_165"  
/note="Organ: testis; Vector: pDNR-LIB; Site 1: SfiI  
(ggcgcataggcc); Site 2: SfiI (ggcgccctggcc); 5' and 3'  
adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CACGGCATTAATGACC-3' and 3' adaptor sequence:  
5'-ATTCTAGAGCCGAGCGGCGGCGCATCATG-dT(30) BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.4 kb  
(range 0.6-3.5 kb). 15/15 colonies contained inserts by  
PCR. This library was enriched for full-length clones and  
was constructed by Clontech Laboratories (Palo Alto, CA).  
Note: this is a NIH-MGC Library."

Query Match 66.8%; Score 694.6; DB 6; Length 733;  
Best Local Similarity 99.4%; Pred. No. 4.1e-146;  
Matches 697; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
ORIGIN  
28 GTGGACCTATTTCCCGGAGACAGGCGTCCAGCTTATGAGTTGTACATATGCGAGGT 87  
Db 2 GGGGACCTATTTCCCGGAGACAGGCGTCCAGCTTATGAGTTGTGTACATATGCGAGGT 61  
Oy 88 AAGAAATGAATGAAGATTATCCGTTAGAAATTCAGAGCTTTTACAGCCCTGGAGACT 147  
Db 62 AAGAAATGAATGAAGATTATCCGTTAGAAATTCAGAGCTTTTACAGCCCTGGAGACT 121  
Oy 148 CCTGGGTGTGTGACGACATGTCTGAGACCATGATGCTGTTTCTAGAAACGAGTTGT 207  
Db 122 CCTGGGTGTGTGACGACATGTCTGAGACCATGATGCTGTTTCTAGAAACGAGTTGT 181  
Oy 208 TGCAGAGTTGACCCATGGAACAAGCAAGAGGTGATTTAGTTTGTGATACACCTTAA 267  
Db 182 TGCAGAGTTGACCCATGGAACAAGCAAGAGGTGATTTAGTTTGTGATACACCTTAA 241  
Oy 268 ATTCATATGTTTGGTTTATTTTGGCACTCAAGAGTTTATCCCAAGAGCATCTAGTA 327  
Db 242 ATTCATATGTTTGGTTTATTTTGGCACTCAAGAGTTTATCCCAAGAGCATCTAGTA 301  
Oy 328 AGCAGGAATCTGAAAGATTCAGAGCTTATGATGATGATGATGATGATGATGATGAT 387  
Db 302 AGCAGGAATCTGAAAGATTCAGAGCTTATGATGATGATGATGATGATGATGATGAT 361  
Oy 388 AGAAGGTGTGCAAGCTGCAAGAGCTGCTGCTTGAATTTGCAAGAGGCACTCTGGG 447  
Db 362 AGAAGGTGTGCAAGCTGCAAGAGCTGCTGCTTGAATTTGCAAGAGGCACTCTGGG 421  
Oy 448 AACCCAAACGAAAAAGCAACCAAAAGTGGCTTAATAAGGAAACCAAAACCTAATCTT 507  
Db 422 AACCCAAACGAAAAAGCAACCAAAAGTGGCTTAATAAGGAAACCAAAACCTAATCTT 481  
Oy 508 TTGTTTGTGATGATCATGTTTCAAAAAGTACATCTTTTATATCATGTTTACATGTAGT 567  
Db 482 TTGTTTGTGATGATCATGTTTCAAAAAGTACATCTTTTATATCATGTTTACATGTAGT 541  
Oy 568 TATGTGACATGTGTGTTTAAATGATTCCTTTTGAATTCATGTATTAATTTACACAT 627  
Db 542 TATGTGACATGTGTGTTTAAATGATTCCTTTTGAATTCATGTATTAATTTACACAT 601  
Oy 628 TACATTTGTATCTGAATCTTTTCTGTGAGAAAGTATGATGTTCTTTGTGATTTT 687  
Db 602 TACATTTGTATCTGAATCTTTTCTGTGAGAAAGTATGATGTTCTTTGTGATTTT 661  
Oy 688 TCATATTAAGCATCATGATGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 728  
Db 662 TCATATTAAGCATCATGATGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 702

RESULT 9  
LOCUS CB574390  
DEFINITION ABEINCOURT\_12795777 NIH MGC\_165 Mus musculus cDNA clone  
IMAGE:30279046 5', mRNA sequence.  
ACCESSION CB574390  
VERSION CB574390.1 GI:29493920  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 731)



AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Dr. Leslie L. Heckert  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: NDCM34 row: 9 column: 23  
 High quality sequence stop: 608.

## FEATURES

source

1..731  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30279046"  
 /issue\_type="primary cultures of Sertoli cells"  
 /lab\_host="DH10B (TI-phage-resistant)"  
 /clone\_lib="NIH\_MGC\_165"  
 /note="Organ: testis; Vector: pDNR-LIB; Site 1: SfiI (ggccatcgcc); Site 2: SfiI (ggccgcctggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCGCATTTGACC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.6-3.5 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA).  
 Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 66.5%; Score 692; DB 6; Length 731;  
 Best Local Similarity 98.6%; Pred. No. 1.6e-145;  
 Matches 699; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

1 CAGAACCGTGTATGCGCTGATCATCTGTCGACACTATTTCCCGAGACAGGCGTCCACG 60  
 24 CAGAACCGTGTATGCGCTGATCATCTGTCGACACTATTTCCCGAGACAGGCGTCCACG 83  
 61 GTATTGAGTTGTCACAAATGCGAGTGAAGAAATGAATGAAGTTATCCCGTAGAAATTC 120  
 84 GTATTGAGTTGTCACAAATGCGAGTGAAGAAATGAAGTTATCCCGTAGAAATTC 143  
 121 ACAGAGCTTTAAACAGCCCTGGAAGAGCTCCCTGGTGTCTGAGACGACATGCTGAAGACA 180  
 144 ACAGAGCTTTAAACAGCCCTGGAAGAGCTCCCTGGTGTCTGAGACGACATGCTGAAGACA 203  
 181 TGATGCTGTTTGAAGACGAGTTGTCAGAAAGTTGACCCATTGGAACAAAGCAAGG 240  
 204 TGATGCTGTTTGAAGACGAGTTGTCAGAAAGTTGACCCATTGGAACAAAGCAAGG 263  
 241 TGGATTAGTTTCTGCTACACCTTAATTCATGTTTGGGTTTATTGGCACTCAAG 300  
 264 TGGATTAGTTTCTGCTACACCTTAATTCATGTTTGGGTTTATTGGCACTCAAG 323  
 301 GAGTTATCCCAAGAGCATCCAGTGAAGAGCAAGTGAAGAAATGAGAGTCTACATGA 360  
 324 GAGTTATCCCAAGAGCATCCAGTGAAGAGCAAGTGAAGAAATGAGAGTCTACATGA 383  
 361 ACAGAGTTAAAGAAATTAACAGACAGAAAGAGCTGCCAAGCTGACAGAGGTGCTGCTT 420  
 384 ACAGAGTTAAAGAAATTAACAGACAGAAAGAGCTGCCAAGCTGACAGAGGTGCTGCTT 443  
 421 CGAGATTGTCAGAAAGCACTCTGGGAAACCCAAAGCAACCAAAAGTGGCTA 480  
 444 CGAGATTGTCAGAAAGCACTCTGGGAAACCCAAAGCAACCAAAAGTGGCTA 503  
 481 ATAAAGGAAGCAACCACTAATCTTTGTTTGTATGATGTTTCAAAAAGTACA 540

Db 504 ATAAAGGAAGCAACCACTAATCTTTGTTTGTATGATGTTTCAAAAAGTACA 563  
 Qy 541 TCCCTTTTAATCACTTACATGATGTTATGACCATGCGTGTAAATGATTCCTT 600  
 Db 564 TCCCTTTTAATCACTTACATGATGTTATGACCATGCGTGTAAATGATTCCTT 623  
 Qy 601 TTGGAATTCATGATTAATTAATTAACATTAATTTGATGATCTGTAATCTTTTGGCTGA 660  
 Db 624 TTGGAATTCATGATTAATTAATTAACATTAATTTGATGATCTGTAATCTTTTGGCTGA 683  
 Qy 661 GAAAGATTAAGTTGCTCTTGTGATTTTCATTAATTAACATCATGATGT 708  
 Db 684 AAAAGATTAAGTTGCTCTTGTGATTTTCATTAATTAACATCATGATGT 731

## RESULT 10

CB598466 715 bp mRNA linear EST 03-APR-2003  
 LOCUS AGENCOURT 12765813 NIH\_MGC\_178 Mus musculus cDNA clone  
 DEFINITION IMAGE:30257772 5', mRNA sequence.

ACCESSION CB598466.1 GI:29516322

VERSION CB598466.1 GI:29516322

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Dr. Michael Brownstein  
 CDNA Library Preparation: Michael Brownstein Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: NDCM59 row: d column: 05  
 High quality sequence stop: 590.

## FEATURES

source

1..715  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30297772"  
 /lab\_host="DH10B (TI-phage-resistant)"  
 /clone\_lib="NIH\_MGC\_178"  
 /note="Organ: lung and heart; Vector: pDNR-LIB; Site 1: SfiI (ggccatcgcc); Site 2: SfiI (ggccgcctggcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-AACGAGTGTATCAAGCAGAGGAGCAATTAACGCGCG-3' and 5'-ATTCTAGAGCCGAGCGCGCCGACATG-dt(30)BN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIH, NIH). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 66.0%; Score 686; DB 6; Length 715;  
 Best Local Similarity 99.3%; Pred. No. 3.6e-144;  
 Matches 689; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 CAGAACCGTGTATGCGCTGATCATCTGTCGACACTATTTCCCGAGACAGGCGTCCACG 60  
 22 CAGAACCGTGTATGCGCTGATCATCTGTCGACACTATTTCCCGAGACAGGCGTCCACG 81  
 61 GTATTGAGTTGTCACAAATGCGAGTGAAGAAATGAATGAAGTTATCCCGTAGAAATTC 120  
 82 GTATTGAGTTGTCACAAATGCGAGTGAAGAAATGAATGAAGTTATCCCGTAGAAATTC 141



FEATURES	LOCATION/Qualifiers
0y	847 GTAAAGATGAGAAATACGATGACTAAATPACCACCTCG-TTTCGTCAGTTAGTTCAA 905
Db	182 GTAAATGATGAGAAATACGATGACTAAATPACCACCTCGTTTCGTCAGTTAGTTCAA 123
0y	906 CATGTTTCGTGATTTTTTTTTTTTTTTAGTAATTCGTCTTGATATTCAAAGT-CAAAA 964
Db	122 CATGTTTCGTGATTTTTTTTTTTTTTTAGAGTAATTCGTCTTGATATTCAAAGTACAAA 63
0y	965 TTGAAACCTTAAGCGCTGACTTTAATTCCTCATGTTCCATTTAAATTAATGTTTCAT 1024
Db	62 TTGAAACCTTAAGCGCTGACTTTAGTTCTTCATGTTCCATTTAAATTAATGTTTCAT 3
0y	1025 TA 1026
Db	2 TA 1
RESULT 12	
LOCUS	BB612968
DEFINITION	BB612968 RIKEN full-length enriched, 10 day neonate skin Mus
ACCESSION	BB612968
VERSION	BB612968.1 GI:15395380
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 685)
AUTHORS	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hizumoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,U., Komno,H., Konda,M., Koya,S., Matsumura,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sobabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.
TITLE	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL	Unpublished (2001)
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.jp, URL: <a href="http://genome.gsc.riken.jp/">http://genome.gsc.riken.jp/</a> Carninci,P., Shibata,Y., Hayasu,N., Sugahara,Y., Shibata,K., Itoh,M., Komno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) Wegli,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanahki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Komno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,T., Shibata,K., Arakawa,T., Ishii,Y. and Hayashizaki,Y. Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp">http://genome.gsc.riken.go.jp</a> ) for further details. e mouse tissues.

[illegible]

Db	542	IGTACTGATGTCACCATGTCGTCGTTTAAATGCAATTCCTTTGGAAATTCATGATATAAATT	601
OY	622	ACACATTCACATTTGTGATACCTGAATCTTTTCTGAGAAAGTAACTGCTTGT	681
Db	602	ACACATTCACATTTGTGATACCTGAATCTTTTCTGAGAAAGTAACTGCTTGT	661
OY	682	TGATTTTCATATTAAGCATCATG	704
Db	662	TGATTTTCATATTAAGCATCATG	684
RESULT 13			
LOCUS	BQ746434		
DEFINITION	UI-M-BRO-bxp-f-16-0-UI.r1 NIH_BMAP_ERO Mus musculus cDNA clone		
ACCESSION	BQ746434		
VERSION	BQ746434.1		
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 690)		
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strusberg, Ph.D.		

Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/INL at:  
<http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)<sup>1</sup>

Seq primer: PYX-5.

### FEATURES

Location/Qualifiers

1. 690  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5711247"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 15.5 dpc"  
/lab\_host="DH10B (11 phage resistant)"  
/clone\_lib="NH BMAP\_E80"  
/note="Organ: Brain; Vector: pUX-Asc; Site\_1: EcoR I;  
Site\_2: Not I; The library was constructed according to  
Bonaldo, Lemon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with EcoR I adaptor, digested with Not I, and then  
cloned directionally into pUX-Asc vector. The library tag  
sequence located between the Not I site and the polyA  
tail, is GTCGCTGGA. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP): 'Gene Discovery in the Developing Mouse Nervous  
System', supported by national institutes of Mental Health  
(NIMH), Hewitt Chin, Ph.D., program coordinator."

## ORIGIN

Query Match	63.6%;	Score 661.6;	DB 5;	Length 690;
Best Local Similarity	99.1%;	Pred. No. 1.1e-138;		
Matches 686; Conservative	0;	Mismatches 4;	Indels 2;	Gaps 2.

57 CACGGTATTGAGTTGGTCAACATGCAGCTGAAGAAATGAATGAAGTTATCCGTGAA 116

Db	1	CACGGTATTGACTTGTCACAAATGGCAGGTGAAAGAAATGAATGAAGATTATCCCGTAGAA	60
Oy	117	ATTCAAGAGTCTTTTAACAGCCCTGGAGAGCTCCCTGGGCTGTGGACAGACATGCTGAAG	176
Db	61	ATTCAAGAGTCTTT-ACAGCCCTGGAGAGCTCCCTGGGCTGTGGACAGACATGCTGAAG	119
Oy	177	ACCATGATGGCTGTTTCTTGAAAAGAGTTGTTGCAGAAATTGGACCCATTGAAACAAGA	236
Db	120	ACCAATGATGGCTGTTTCTTGAAAACAGTGTGTCAGAAATTGGACCCATTGAAACAAGA	179
Oy	237	AAGGTGATTTAGTTGTTGTCATPACCTTAAATTCGAATGTTTGGGTTTATTGGCAACT	296
Db	180	AAGGTGATTTAGTTTCTGCATPACCTTAAATTCGAATGTTTGGGTTTATTGGCAACT	239
Oy	297	CAAGGATTTATCCCAAAGAGCATCTCAGTGAAGCAGAACTGGAAAGAAATCAGAGTCTAC	356
Db	240	CAAGGATTTATCCCAAAGAGCATCTCAGTGAAGCAGAACTGGAAAGAAATCAGAGTCTAC	299
Oy	357	ATGAACAAGTTAAAGAAATACAGACAAAGAAAGGCTGCCAAGTGGACAGAGTGTCT	416
Db	300	ATGAACAAGTTAAAGAAATACAGACAAAGAAAGGCTGCCAAGTGGACAGAGTGTCT	359
Oy	417	GCTTCGAGATTGTCAGAGAAGCACTCTGGGAAACCCAAACGAAAAAGACACCAAAAGTG	476
Db	360	GCTTCGAGATTGTCAGAGAAGCACTCTGGGAAACCCAAAGCAAAAAGCACACAAAGTG	419
Oy	477	GCTAATAAAGGAAAAAGCAACCTAACTTTGGTTTGTATGTACATGTTTCAAAG	536
Db	420	GCTAATAAAGGAAAAAGCAACCTAACT-TTTGGTTTGTATGTACATGTTTCAAAG	478
Oy	537	TACATCCTTTTAAATCAGTTTACATGTAGTATGTGACATGTGTGTATTAAATGAGATT	596
Db	479	TACATCCTTTTAAATCAGTTTACATGTATGTATGTGACATGTGTGTATTAAATGAGATT	538
Oy	597	CCTTTTGGAAATTCATGTATTAATTTACACATTACATTTGTGTATCTGAATCTTTTTTTTG	656
Db	539	CCTTTTGGAAATTCATGTATTAATTTACACATTACATTTGTGTATCTGAATCTTTTTTTTG	598
Oy	657	CTGAGAAAGATTAAGTGTCTTTTGTGTGATTTTCAATATAAAGCATCATGATGTGTTAATA	716
Db	599	CTGAGAAAGATTAAGTGTCTTTTGTGTGATTTTCAATATAAAGCATCATGATGTGTTAATA	658
Oy	717	TTGTATAGATATTCATAAAGCAGTTGTGAATC 748	
Db	659	TTGTATAGATATTCATAAAGCAGTTGTGAATC 690	

RESULT 14					
LOCUS	CD352734				
DEFINITION	CD352734	671 bp	mRNA	linear	EST 15-JUL-2003
ACCESSION	U1-M-0L0-cfy-h-18-0-UI.r1 NIH BMAP_G10				
VERSION	IMAG3:30359321.5				
KEYWORDS	CD352734				
SOURCE	EST.				
ORGANISM	Mus musculus (house mouse)				
	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 671)				
AUTHORS	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D.				

Tissue Procurement: Dr. Jim Jahn, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mouse1.html>  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES Seq primer: PYX-5.  
Location/Qualifiers  
source 1..671  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/accession="U00000"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30359321"  
/tissue\_type="Whole brain"  
/dev\_stage="1, 5 and 15 days newborn"  
/lab\_host="DH10B (TI phage resistant)"  
/clone\_lib="NIH BMAP GL0"  
/note="Organ: Brain; Vector: PYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bontad, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction. ligated with EcoR I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTCAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN  
Query Match 62.9%; Score 654.2; DB 6; Length 671;  
Best Local Similarity 99.4%; Pred. No. 5.4e-137;  
Matches 667; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
74 CACAATGGCAGGTGAAGAAATGAATGAATTC-CGTGAATTCACGAGCTTTAA 132  
1 CACAATGGCAGGTGAAGAAATGAATGAATTC-CGTGAATTCACGAGCTTTAA 60  
133 CACCCCTGGAGACTCCCTGGGCTGTGACGACGACGCTGGAAGACATGATGCTTT 192  
61 CACCCCTGGAGACTCCCTGGGCTGTGACGACGACGCTGGAAGACATGATGCTTT 120  
193 CTGAAACGAGTGTGACAGATGGACCATTTGAAACGAGCAAGATGATGATTT 252  
121 CTGAAACGAGTGTGACAGATGGACCATTTGAAACGAGCAAGATGATGATTT 180  
253 CTGCAATACACTTAATTCATGTTTGGTTTATTTGGCACTCAAGATTAATCCA 312  
181 CTGCAATACACTTAATTCATGTTTGGTTTATTTGGCACTCAAGATTAATCCA 240  
313 AAGAGCATTCAGTGAAGCAGAACTGGAAGATCAGAGTTCATGAAGATTAAG 372  
241 AAGAGCATTCAGTGAAGCAGAACTGGAAGATCAGAGTTCATGAAGATTAAG 300  
373 AATTAACAGACAGAAAGAGGCTGCCAAGCTGGACAGAGGCTGCTGAGATTTGCA 432  
301 AATTAACAGACAGAAAGAGGCTGCCAAGCTGGACAGAGGCTGCTGAGATTTGCA 360  
433 AGAAGCACTCTGGAAACCCAAAGCAAAAGCAACCAAAAGTGGCTTAAGAGGAAA 492  
361 AGAAGCACTCTGGAAACCCAAAGCAAAAGCAACCAAAAGTGGCTTAAGAGGAAA 420  
493 GCAGAACTTAATCTTTGGTTGATGATGATGATGATGATGATGATGATGATGAT 552  
421 GCAGAACTTAATCTTTGGTTGATGATGATGATGATGATGATGATGATGATGAT 480  
553 AGTTTCAATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 612  
481 AGTTTCAATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 540  
613 TATTAATTTACATTAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 672  
541 TATTAATTTACATTAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 600

ORIGIN  
Query Match 62.7%; Score 651.6; DB 6; Length 772;  
Best Local Similarity 98.4%; Pred. No. 2.1e-136;  
Matches 679; Conservative 0; Mismatches 9; Indels 2; Gaps 2;  
673 TGTCTTGTGATTTTCATATAAGCATGATGATGATGATGATGATGATGATGAT 732  
601 TGTCTTGTGATTTTCATATAAGCATGATGATGATGATGATGATGATGATGAT 660  
733 AAGCAGTTGTG 743  
661 AAGCAGTTGTG 671  
RESULT 15  
CB320669 772 bp mRNA linear EST 04-MAR-2003  
LOCUS AGENCOURT 12427115 NIH\_MGC\_165 Mus musculus cDNA clone  
DEFINITION IMAGE:30285684 5', mRNA sequence.  
ACCESSION CB320669  
VERSION CB320669.1 GI:28844904  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 772)  
NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Straube, Ph.D.  
Email: cga@nci.nih.gov  
Tissue Procurement: Dr. Leslie L. Heckert  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDCM51 row: 1 column: 13  
High quality sequence stop: 578.  
location/Qualifiers  
1..772  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30285684"  
/tissue\_type="primary cultures of Sertoli cells"  
/lab\_host="DH10B (TI-phage-resistant)"  
/clone\_lib="NIH\_MGC\_165"  
/note="Organ: testis; Vector: pDNR-LIB; Site 1: SfiI (ggccatcgcc); Site 2: SfiI (ggccgctcgcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCATTTATGCG-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGAGCGCGCATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.6-3.5 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC library."

ORIGIN  
Query Match 62.7%; Score 651.6; DB 6; Length 772;  
Best Local Similarity 98.4%; Pred. No. 2.1e-136;  
Matches 679; Conservative 0; Mismatches 9; Indels 2; Gaps 2;  
1 CAGAGCCGTGTCATGCGCTCATGTCGACCTAATTTCCGAGACAGCGCTCCACG 60  
21 CAGAGCCGTGTCATGCGCTCATGTCGACCTAATTTCCGAGACAGCGCTCCACG 80  
61 GTATTAGTTGTGCAAAAGGAGGAGAAATGAATGAATTAATCCCGTGAATTC 120  
81 GTATTAGTTGTGCAAAAGGAGGAGAAATGAATGAATTAATCCCGTGAATTC 140  
121 ACAGAGCTTTAAACAGCCCTGAGAGCTCCCTGGGCTGTGAGCAGATGTAAGCA 180  
141 ACAGAGCTTTAAACAGCCCTGAGAGCTCCCTGGGCTGTGAGTGAATGTAAGCA 200

```
OY 181 TGATGGCTGTTCTAGAAAGAGTTGTGCAAGTTGACCCATTTGAACTAAGCAAAGG 240
    |||||
DB 201 TGATGGCTGTTCTAGAAATGATGTGTGCAAGAGTTGACCCATTTGAACTAAGCAAAGG 260
    |||||
OY 241 TGGATTATGTTTGTGCATACACTTAATTCATATGTTTGGGTTTATTTGGCAACTCAAG 300
    |||||
DB 261 TGGATTATGTTTGTGCATACACTTAATTCATATGTTTGGGTTTATTTGGCAACTCAAG 320
    |||||
OY 301 GAGTTAATCCCAAGAGCATCCATGAAAGCAAGAACTGAAAGAAATCAGAGTCTACATGA 360
    |||||
DB 321 GAGTTAATCCCAAGAGCATCCATGAAAGCAAGAACTGAAAGAAATCAGAGTCTACATGA 380
    |||||
OY 361 ACAGAGTTAAAGAAATTAACAGACAAAGAAAGGCTGCCAAGCTGSAACAGAGTGTGCTT 420
    |||||
DB 381 ACAGAGTTAAAGAAATTAACAGACAAAGAAAGGCTGCCAAGCTGSAACAGAGTGTGCTT 440
    |||||
OY 421 CGAGATTGTCAAGAAAGGCACTGTGGAACTCAAGCAAAAGCAACACCAAAAGTGCTTA 480
    |||||
DB 441 CGAGATTGTCAAGAAAGGCACTGTGGAACTCAAGCAAAAGCAACACCAAAAGTGCTTA 500
    |||||
OY 481 ATAAAGGAAAGCAAACTAATCTTTGGTTTGAATGATCAATGTTTCAAAAAGTACA 540
    |||||
DB 501 ATAAAGGAAAGCAAACTAATCTTTGGTTTGAATGATCAATGTTTCAAAAAGTACA 560
    |||||
OY 541 TCCTTTTATCGATTACAGTGTATGTGACCATGTGGTTTAATGATTCCTT 600
    |||||
DB 561 TCCTTTTATCGATTACAGTGTATGTGACCATGTGGTTTAATGATTCCTT 620
    |||||
OY 601 TT-GGAATTCATGTATTAATTTACACATTAATTTGTGTGATCTGAATC-TTTTTTGTGCT 658
    |||||
DB 621 TTGGGATTCATGTATTAATTTACACATTAATTTGTGTGATCTGAATC-TTTTTTGTGCT 680
    |||||
OY 659 GAGAAAGATTAAAGTTGTCTTTGTGATTTT 688
    |||||
DB 681 GAAAAAGATTAAAGTTGTCTTTGTGATTTT 710
    |||||
```

Search completed: January 4, 2005, 13:36:35  
Job time : 3838.49 secs





QY	61	GTATTGAGTTGGTCA	CAA	TGG	CAG	TGA	AGAA	TGA	ATGA	ATTA	TCC	GTAGA	AA	TT	C	120											
Db	61	GTATTGAGTTGGTCA	CA	ATGG	CAG	TGA	AGAA	TGA	ATGA	ATTA	TCC	GTAGA	AA	TT	C	120											
QY	121	ACGAGTCTTTAA	C	AGCC	CTG	AG	AG	CTCC	CTG	GGT	GTG	AG	CA	CA	TG	CTGA	AG	CCA	180								
Db	121	ACGAGTCTTTAA	C	AGCC	CTG	AG	AG	CTCC	CTG	GGT	GTG	AG	CA	CA	TG	CTGA	AG	CCA	180								
QY	181	TGATGGCTGTTT	CTA	GA	AC	GA	GT	TTG	TG	CA	AA	GT	TGA	CC	CA	TTG	GA	CA	AG	CAA	240						
Db	181	TGATGGCTGTTT	CTA	GA	AC	GA	GT	TTG	TG	CA	AA	GT	TGA	CC	CA	TTG	GA	CA	AG	CAA	240						
QY	241	TGATTTAGTTT	CTG	CA	TAC	CTTAA	TT	CA	AT	TTG	GG	TTT	ATT	TG	GA	CA	CT	CA	AG	300							
Db	241	TGATTTAGTTT	CTG	CA	TAC	CTTAA	TT	CA	AT	TTG	GG	TTT	ATT	TG	GA	CA	CT	CA	AG	300							
QY	301	GAGTAAATCC	CA	AA	AG	CA	CT	CA	GTGA	AG	CA	GGAA	CTG	MA	AGAA	T	CA	GA	GT	CA	TA	GA	360				
Db	301	GAGTAAATCC	CA	AA	AG	CA	CT	CA	GTGA	AG	CA	GGAA	CTG	MA	AGAA	T	CA	GA	GT	CA	TA	GA	360				
QY	361	ACAGATTTAA	GA	AA	ATTA	C	GA	CA	AA	GA	AG	CTCC	AA	GTG	CA	CA	AG	CTG	CT	420							
Db	361	ACAGATTTAA	GA	AA	ATTA	C	GA	CA	AA	GA	AG	CTCC	AA	GTG	CA	CA	AG	CTG	CT	420							
QY	421	CGAGATTTGT	CA	AA	AG	CA	CT	CTG	GGAA	CC	CA	AA	GA	CA	CA	CA	AA	AG	GTG	CT	480						
Db	421	CGAGATTTGT	CA	AA	AG	CA	CT	CTG	GGAA	CC	CA	AA	GA	CA	CA	CA	AA	AG	GTG	CT	480						
QY	481	ATTAAGGGA	AA	AA	GA	CA	CA	CT	ATCT	TT	GG	TTT	GA	GT	A	AT	GT	TT	CA	AA	AG	T	GA	540			
Db	481	ATTAAGGGA	AA	AA	GA	CA	CA	CT	ATCT	TT	GG	TTT	GA	GT	A	AT	GT	TT	CA	AA	AG	T	GA	540			
QY	541	TCCTTTTAA	T	CA	GT	T	CA	AT	GA	TA	GT	GA	CA	TA	GT	GA	CA	TA	GT	GA	CA	TA	GT	GA	540		
Db	541	TCCTTTTAA	T	CA	GT	T	CA	AT	GA	TA	GT	GA	CA	TA	GT	GA	CA	TA	GT	GA	CA	TA	GT	GA	540		
QY	601	TTGGAATTC	AT	CA	GT	TA	AA	TT	T	CA	CA	TT	A	CA	TT	T	GA	CA	TT	T	GA	CA	TT	T	GA	600	
Db	601	TTGGAATTC	AT	CA	GT	TA	AA	TT	T	CA	CA	TT	A	CA	TT	T	GA	CA	TT	T	GA	CA	TT	T	GA	600	
QY	661	GAAATTTA	AG	T	GT	CT	TT	T	GT	GA	TTT	CA	T	A	T	A	A	A	G	A	G	A	G	A	G	A	660
Db	661	GAAATTTA	AG	T	GT	CT	TT	T	GT	GA	TTT	CA	T	A	T	A	A	A	G	A	G	A	G	A	G	A	660
QY	721	AAGATATTC	T	CA	TA	AG	CA	GT	T	GA	AA	T	CT	CT	GA	TA	AA	AG	CA	T	GA	TA	AA	T	CT	CT	720
Db	721	AAGATATTC	T	CA	TA	AG	CA	GT	T	GA	AA	T	CT	CT	GA	TA	AA	AG	CA	T	GA	TA	AA	T	CT	CT	720
QY	781	AAATGA	AA	CA	AA	GA	TA	TT	T	GA	AG	TG	CT	GA	T	CT	GA	T	GA	AG	CA	TT	T	T	T	T	780
Db	781	AAATGA	AA	CA	AA	GA	TA	TT	T	GA	AG	TG	CT	GA	T	CT	GA	T	GA	AG	CA	TT	T	T	T	T	780
QY	841	TTTTCA	TA	AT	GA	T	GA	GA	AA	TA																	

DEFINITION	Sequence 3 from Patent WO963071.
ACCESSION	AX009363
VERSION	AX009363.1
KEYWORDS	GI:996664
SOURCE	Mus sp.
ORGANISM	Mus sp.
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1
TITLE	Rochbarth, K., Werner, D. and Stammer, H.
JOURNAL	Method for triggering apoptosis in cells
Patent:	WO 96/3071-A 3 09-DEC-1999;
DEUTSCHES KEEBESORSCH (DE);	
STAMMER HERMANN (DE);	
WERNER DIETER	
FEATURES	location/Qualifiers
source	1..1040
CDS	/organism="Mus sp." /mol_type="unassigned DNA" /db_xref="taxon:10095" 78..503 /note="unnamed protein product" /codon_start=1 /protein_id="CAC07451.1" /db_xref="GI:996665" /translation="MAGEWMEDYVEIHESITALESLGAVDMLKTMVAVSRNELL OKUDLEQAKVDLSAYATLNSMFVWYLAITGVNPEHVKQELERIRVYNRVEITL KKKAAKLDGSAISRFEVKALWEPKRSKTPKANKSGH"
ORIGIN	
Query Match	100.0%; Score 1040; DB 6; Length 1040;
Best Local Similarity	100.0%; Pred. No. 3.2e-197;
Matches 1040;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 CCAAGCCGTGTCATGCGCTCATCTGTCGACCTATTTCCCGAGACAGCGCTCACG 60
DB	1 CCAAGCCGTGTCATGCGCTCATCTGTCGACCTATTTCCCGAGACAGCGCTCACG 60
QY	61 GATTTGAGTTGGTGCACAAATGGAGGTGAAAGAAATGAATGAATTTATCCGTAGAAATTC 120
DB	61 GATTTGAGTTGGTGCACAAATGGAGGTGAAAGAAATGAATGAATTTATCCGTAGAAATTC 120
QY	121 ACGAGCTTTTAAACAGCCCTGGAAGCTCCCTGGGTGCTGTGACGACATGCTGAAGCA 180
DB	121 ACGAGCTTTTAAACAGCCCTGGAAGCTCCCTGGGTGCTGTGACGACATGCTGAAGCA 180
QY	181 TGAATGCGTGTTCAGAAACGAGTTGTTGCGAAGTTGACCCATTTGGAACAGCAAGG 240
DB	181 TGAATGCGTGTTCAGAAACGAGTTGTTGCGAAGTTGACCCATTTGGAACAGCAAGG 240
QY	241 TGGATTTAGTTTCGCATACACCTTAAATTCATGTTTGGGTTTATTTGGCAACTCAAG 300
DB	241 TGGATTTAGTTTCGCATACACCTTAAATTCATGTTTGGGTTTATTTGGCAACTCAAG 300
QY	301 GAGTTAATCCCAAGAGATCCAGTGAAGAGAACTGGAAGAAATACAGCTTACATGA 360
DB	301 GAGTTAATCCCAAGAGATCCAGTGAAGAGAACTGGAAGAAATACAGCTTACATGA 360
QY	361 ACAAGATTAAAGAAATTAACAGACAGAAAGAGCTGCCAAGCTGACAGAGGTGCTT 420
DB	361 ACAAGATTAAAGAAATTAACAGACAGAAAGAGCTGCCAAGCTGACAGAGGTGCTT 420
QY	421 CGAATTTGTCAGAGAGGCACTCTGGGAACCCGAAAGGAAAGACACCAAAAGTGGCTA 480
DB	421 CGAATTTGTCAGAGAGGCACTCTGGGAACCCGAAAGGAAAGACACCAAAAGTGGCTA 480
QY	481 ATAAAGGAAAGCAACATATCTTTTGGTTTGAATGATGTTTCAAAAAGTACA 540
DB	481 ATAAAGGAAAGCAACATATCTTTTGGTTTGAATGATGTTTCAAAAAGTACA 540
QY	541 TCCGTTTATCGTTTACATGCTACTATTTGACACCATGTGTTTAAATGATTCCTT 600
DB	541 TCCGTTTATCGTTTACATGCTACTATTTGACACCATGTGTTTAAATGATTCCTT 600





QY 841 TTTTCAGTATGATGAGAAATACAGTACTTAATACCCACTGTCTTCTGTTCAGTTAG 900  
Db 841 TTTTCAGTATGATGAGAAATACAGTACTTAATACCCACTGTCTTCTGTTCAGTTAG 900  
QY 901 TTCAACATGTTTCTGATTTTTTTTTTTTTTTGAGTAACTTCTGTTCAGTTAGTTCAAAAGTC 960  
Db 901 TTCAACATGTTTCTGATTTTTTTTTTTTTTTGAGTAACTTCTGTTCAGTTAGTTCAAAAGTC 960  
QY 961 AAAATGAAACCTTAAGCTGTACTTTAATCTTCATGTTCCATTTAAATAAATGTTTC 1020  
Db 961 AAAATGAAACCTTAAGCTGTACTTTAATCTTCATGTTCCATTTAAATAAATGTTTC 1020  
QY 1021 TCATTAACTGTATGAGAAA 1040  
Db 1021 TCATTAACTGTATGAGAAA 1040

RESULT 4  
LOCUS AF031426 1040 bp DNA linear ROD 25-DEC-1997  
DEFINITION Mus musculus small unique nuclear receptor co-repressor (SUN-COR)  
ACCESSION AF031426  
VERSION AF031426.1 GI:2642585  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM  
REFERENCE 1. Bases 1 to 1040  
AUTHORS Lazar, M.A., Dawson, J., Lavinsky, R.M., Glass, C.K., Rosenfeld, M.G. and Zamir, I. and Lazar, M.A.  
TITLE Cloning and characterization of a corepressor and potential component of the nuclear hormone receptor repression complex  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (26), 14400-14405 (1997)  
MEDLINE 9405624  
PUBMED 2 (bases 1 to 1040)  
REFERENCE Zamir, I. and Lazar, M.A.  
AUTHORS Direct Submission  
TITLE Submitted (24-OCT-1997) Medicine, University of Pennsylvania, 611 Curie Blvd., Philadelphia, PA 19104, USA  
JOURNAL CRB, 415  
FEATURES  
source location/Qualifiers  
1..1040  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/note="Induced during adipocyte differentiation of 3T3-L1 cells and during myogenic differentiation of C2-C12 cells"  
gene /gene="SUN-COR"  
80..505  
/gene="SUN-COR"  
/note="corepressor for nuclear hormone receptors; CID"  
/codon\_start=1  
/product="small unique nuclear receptor co-repressor"  
/protein\_id="AAC53520.1"  
/db\_xref="GI:2642586"  
/translation="MAGEKNEDEYPIEIHESLTALFSSLGAVDMLKTMVAVSNNEL  
OKDPLEQAVDVSVATLNSMFVYLATGVNPKSHPYKQELERIRVYMNVRKETTD  
KKAAKXKDRGAASRFVKNALMEPKASTPRVANKRSKH"

ORIGIN

Query Match 92.3%; Score 960.4; DB 10; Length 1040;  
Best Local Similarity 97.9%; Pred. No.2.4e-181;  
Matches 1016; Conservative 0; Mismatches 16; Indels 6; Gaps 4;

QY 1 CAGAAGCGCTGTATGCGGTATCATCTTATTTCCCGAGACAGGCGTCCAG 60  
Db 3 CAGAAGCGCTGTATGCGGTATCATCTTATTTCCCGAGACAGGCGTCCAG 62  
QY 61 GTATTGAGTTGTCAATGCGAGGTGAAGAAATGATTAAGATTATCCCGTGAATTC 120

Db 63 GTATTGAGTTGTCAATGCGAGGTGAAGAAATGATTAAGATTATCCCGTGAATTC 122  
QY 121 ACGAGCTTTTAAACAGCCCTGGAGAGCTCCCGGGTGTGTGAGACACATGCTGAAGCCA 180  
Db 123 ACGAGCTTTTAAACAGCCCTGGAGAGCTCCCGGGTGTGTGAGACATGCTGAAGCCA 182  
QY 181 TGATGCGCTGTTCTTGAAACGAGTTGTGCAAGTTGGAACCCATTGGAACGAAG 240  
Db 183 TGATGCGCTGTTCTTGAAATGAGTTGTGCAAGTTGGAACCCATTGGAACGAAG 242  
QY 241 TGATTTAGTTTCTGCAATACCTTAAATTCANGTGTTGGTTTATTTGGCACTCAG 300  
Db 243 TGATTTAGTTTCTGCAATACCTTAAATTCANGTGTTGGTTTATTTGGCACTCAG 302  
QY 301 GAGTTAATCCCAAGAGCATTCAGTGAAGGAGCACTGGAAGATTCAGAGTTCATGA 360  
Db 303 GAGTTAATCCCAAGAGCATTCAGTGAAGGAGCACTGGAAGATTCAGAGTTCATGA 362  
QY 361 ACAAGTTTAAAGAAATPACAGACAGAGAGAGGCTGCCAAGCTGGAAGAGTGTCTT 420  
Db 363 ACAAGTTTAAAGAAATPACAGACAGAGAGAGGCTGCCAAGCTGGAAGAGTGTCTT 422  
QY 421 CGAGATTGTCAAGAAAGCACTCTGGGAAACCCAAAGCAAAAGCAACCAAAAGGCTA 480  
Db 423 CGAGATTGTCAAGAAAGCACTCTGGGAAACCCAAAGCAAAAGCAACCAAAAGGCTA 482  
QY 481 ATTAAGGAAAGCAAAAGCACTAATCTTTGGTTTGAATGATATGTTTCAAAAAGTCA 540  
Db 483 ATTAAGGAAAGCAAAAGCACTAATCTTTGGTTTGAATGATATGTTTCAAAAAGTCA 542  
QY 541 TCTTTTAAATCACTTAAATGATGATGATGACCATGTGGTGTAAAGGATTCCT 600  
Db 543 TCTTTTAAATCACTTAAATGATGATGATGACCATGTGGTGTAAAGGATTCCT 602  
QY 601 TTGAATTCATGATTAATTAATTAACATTAACATTTGTGATGATGATGATGATGAT 660  
Db 603 TTGAATTCATGATTAATTAATTAACATTAACATTTGTGATGATGATGATGATGAT 662  
QY 661 GAAAGATTAAAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 720  
Db 663 GAAAGATTAAAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 722  
QY 721 AAGATTCTATTAAGCACTGTTGAAATCAATGTTCTGTGTAACATTTGTAGTGTG 780  
Db 723 AAGATTCTATTAAGCACTGTTGAAATCAATGTTCTGTGTAACATTTGTAGTGTG 782  
QY 781 AAGTGAACATGATTAATTAAGAGTGTGTAACCTCGAGGCTGAAGCACTTTG 840  
Db 783 AAGTGAACATGATTAATTAAGAGTGTGTAACCTCGAGGCTGAAGCACTTTG 841  
QY 841 TTTTCAGTATGATGAGAAATACAGTACTTAATACCCACTGTG-TTTCGTTCAGTTA 899  
Db 842 TTTTCAGTATGATGAGAAATACAGTACTTAATACCCACTGTGTTCTGTTCAGTTA 901  
QY 900 GTTCAACATGTTTCTGATTTTTTTTTTTTTTTGAG---TAAATCTGTCTGATATTC 956  
Db 902 GTTCAACATGTTTCTGATTTTTTTTTTTTTTTGAGTAAATGCTGTGATATTC 961  
QY 957 AGT-CAAAATTTGAACCTTAAGGCTGATCTTAATCTTCATGTTCCATTTAAATAA 1015  
Db 962 AGTCAAAATTTGAACCTTAAGGCTGATCTTAATCTTCATGTTCCATTTAAATAA 1021

RESULT 5  
LOCUS BC005436 2000 bp mRNA linear ROD 29-JUN-2004  
DEFINITION Mus musculus nuclear DNA binding protein, mRNA (cDNA clone MGC:5983 IMAGE:359182), complete cds.

ACCESSION BC005436  
VERSION GI:13529391  
KEYWORDS MGC  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 2000)  
Strausberg, R.L., Peltingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenman, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullihy, S.J., Bosak, S.A., McEwan, P.J., Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fehey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A., Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 2000)  
Strausberg, R.  
Direct Submission  
Submitted (27-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [gcgaps-f@mail.nih.gov](mailto:gcgaps-f@mail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAX Plate: 11 Row: d Column: 7  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10101123.  
Location/Qualifiers  
1. 2000  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="MGC:5983 IMAGE:3591682"  
/tissue\_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clone1) outgrowth infected with the virus MMTV."  
/clone\_lib="NCI CGAP Lu29"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPORT6"  
1. 2000  
/gene="Cld"  
/note="synonym: SUN-COR"  
/db\_xref="LocusID:57316"  
/db\_xref="MGI:1927354"

CDS  
47..472  
/gene="Cld"  
/codon\_start=-1  
/product="nuclear DNA-binding protein"  
/protein\_id="AAH05436.1"  
/db\_xref="GI:13529392"  
/db\_xref="LocusID:57316"  
/db\_xref="MGI:1927354"  
/translation="MAGEKINEDYPIEIHSLRALBESSLCAVDMLKTMVAVERNEL  
QKDPBQAKVDVLSATTLNLSMFVVIATGVKPKHVPQELERLRVYNNRVEITD  
KKAAKLRGAASRPVKNALMBEPKASTPVANKSKKH"  
ORIGIN  
Query Match 90.8%; Score 944.6; DB 10; Length 2000;  
Best Local Similarity 97.8%; Pred. No. 3,1e-178;  
Matches 989; Conservative 0; Mismatches 19; Indels 3; Gaps 3;  
32 GACCTATTTCCGGAGACAGAGCGCTCCACGGTATTAAGTGGTCAATGCGAGGTGAAGA 91  
1 GACCTATTTCCGGAGACAGAGCGCTCCACGGTATTAAGTGGTCAATGCGAGGTGAAGA 60  
92 AATGAATGAAGATTATTCGCTAGAAATTCACGAGCTTTTAACAGCCTGAGAGCTCCCT 151  
61 AATGAATGAAGATTATTCGCTAGAAATTCACGAGCTTTTAACAGCCTGAGAGCTCCCT 120  
152 GGGTGGCTGTGAGACGACATGCTGGAAGACATGAGTGGCTTTCTTGAAGAAGAGTTGGCA 211  
121 GGGTGGCTGTGAGACATGCTGGAAGACATGAGTGGCTTTCTTGAAGAAGAGTTGGCA 180  
212 GAAGTTGACCCATTTGGAACAAGAGTGAATTTAGTTTTCATACACCTTAATTC 271  
181 GAAGTTGACCCATTTGGAACAAGAGTGAATTTAGTTTTCATACACCTTAATTC 240  
272 AATGTTTGGGTTTATTTGGCAATCTCAAGAGTTAATCCCAAGAGATCAAGTGAAGCA 331  
241 AATGTTTGGGTTTATTTGGCAATCTCAAGAGTTAATCCCAAGAGATCAAGTGAAGCA 300  
332 GGAATGGAAGAATCAAGATCTACATGAACAAGTTAAAGAAATACACACAAGAA 391  
301 GGAATGGAAGAATCAAGATCTACATGAACAAGTTAAAGAAATACACACAAGAA 360  
392 GGGTGGCAAGTGAAGAGAGTGGCTGCTGAGATTTGTCAGAAGCACTGGGAAC 451  
361 GGGTGGCAAGTGAAGAGAGTGGCTGCTGAGATTTGTCAGAAGCACTGGGAAC 420  
452 CAAGCAAGAAAGACACCAAAAGTGGCTAATAAGGAAAGCAACATTAATCTTTGG 511  
421 CAAGCAAGAAAGACACCAAAAGTGGCTAATAAGGAAAGCAACATTAATCTTTGG 480  
512 TTTTGAATGATCATGTTTCAAAAAGTACATCTTTTAAATCAATGATGATGATG 571  
481 TTTTGAATGATCATGTTTCAAAAAGTACATCTTTTAAATCAATGATGATGATG 540  
572 TGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 631  
541 TGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
632 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 691  
601 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660  
692 AATAAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 751  
661 AATAAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720  
752 ATGTTCTGTGTAACAATTTGATGATGATGATGATGATGATGATGATGATGATG 811  
721 ATGTTCTGTGTAACAATTTGATGATGATGATGATGATGATGATGATGATGATG 780  
812 CTGTAGACCTCGAGGTGAAGACATTTGTTTCAATGATGATGATGATGATGATG 871  
781 CTGTAGACCTCGAGGTGAAGACATTTGTTTCAATGATGATGATGATGATGATG 839

```

QY 872 AATACCACTCTG-TTCTGTTCACTAGTCAACAGTTTCGATTTTTTTTTTTT 930
DB 840 AATACCACTCTGTTCTGTTCTGTTCAACAGTTTCGATTTTTTTTTTTT 899
QY 931 TTGAGTAATTCGTTCTGTTCAATTCAGT-CAGAAATGAACCTTAAAGCGTCTTAA 989
DB 900 GGAGTAATTCGTTCTGTTCAATTCAGTCAACAAATTCAGCTTAAAGCGTCTTAA 959
QY 990 TTCTTCATGTTCCATTTAAATTAATTAATTCCTCACTTAAGTGAAGAAA 1040
DB 960 TTCTTCATGTTCCATTTAAATTAATTAATTCCTCACTTAAGTGAAGAAA 1010

RESULT 6
AC109025/c
LOCUS
DEFINITION
AC109025 215616 bp DNA linear HTG 20-NOV-2002
Rattus norvegicus clone CH230-276A24, *** SEQUENCING IN PROGRESS
***
AC109025
AC109025.5 GI:25139227
HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 215616)
Mizny,D,Marie, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alstbrooks,S, Amlin,A, Anguiano,D,
Alyababehchi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Bernhamed,F,
Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryant,N, Buhay,C, Burch,P, Butrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, P.Souza,L,
Davila,M,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K,
Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
Egan,A, Escotto,M, Eugene,C, Evans,C, Falla,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Frazer,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
Gebrgeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guvvara,W,
Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K,
Harvey,Y, Havlik,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M,
Hollins,B, Howells,S, Hulik,S, Hume,J, Idelbird,D, Jackson,A,
Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorenshewa,L, Louissege,H, Lozano,R,J, Lu,X, Ma,J,
Maheshwari,M, Mahindartine,M, Mahmood,M, Malloy,K, Mangum,A,
Mangum,B, Mapua,P, Martin,K, Martin,R, Martinec,E,
Mawhinney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
Morgan,M, Morris,K, Morris,S, Muidasa,M, Murphy,M, Nair,L,
Nankervis,C, Neal,D, Newton,G, Olarunpusoon,A, Pal,S, Paks,K,
Pasternak,S, Paul,H, Perez,A, Perez,L, Pfankoch,C,
Plopper,F, Polndexter,A, Popovic,D, Primus,E, Fu,L, Li,
Piazto,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,
Riley,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
Sanders,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J,
Shetty,U, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smjs,D,
Sneale,M, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J,
Steinle,M, Strong,R, Sutton,A, Svatek,A, Tabor,P, Taylor,C,
Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Umani,K,
Valas,R, Vera,V, Villaseana,D, Waldron,L, Walker,B, Wang,J,
Wang,O, Wang,S, Warren,J, Warren,R, Wei,X, White,F,
Williams,G, Willson,R, Wleczky,R, Woodem,H, Worley,K,
Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V,

```

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Mederhauern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.  
Unpublished  
Direct Submission  
2 (bases 1 to 215616)  
Worley, K. C.  
Submitted (03-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 215616)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 20, 2002 this sequence version replaced gi:23603213.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both ends and sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

## COMMENT

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GPRH  
Center clone name: CH230-276A24  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 181529 bases at least Q40  
Consensus quality: 183677 bases at least Q30  
Consensus quality: 185417 bases at least Q20  
Estimated insert size: 187302; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the subcloner.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 215616: contig of 215616 bp in length.  
Location/Qualifiers  
1. 215616  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-276A24"  
1. 1067  
/note="wgs end\_extension  
clone end:5g6"  
8625..10234  
/note="wgs end\_extension  
clone end:5p6"  
11405..12230  
/note="clone boundary  
clone\_end:5p6"

## FEATURES

## source

## misc\_feature

## misc\_feature

## misc\_feature

site:  
end\_sequence: BZ161264"  
misc\_feature  
149323..150442  
/note="wgs\_contig"  
complement(212205..213062)  
/note="clone\_boundary"  
clone\_end: 17  
site:  
end\_sequence: BZ161263"

## ORIGIN

Query Match 73.8%; Score 767.8; DB 2; Length 215616;  
Best Local Similarity 89.0%; Pred. No. 3.2e-143;  
Matches 932; Conservative 0; Mismatches 102; Indels 13; Gaps 9;

QY 1 CAGAGCCGTGTCATGCGTCATCATCGTCGACCTATTTCCCGAGACAGCGCTCCACG 60  
DB 181953 CAAAAACCGTCATGCGTCATCATCGTCGACCTCTTTCCGAGACAGCGCTCCACG 181894  
QY 61 GTATTGAGTGTGCACAAATGCGAGGTGAAGAAATGAAGATTATTCCTGTAATAATTC 120  
DB 181893 GTATTGAGTGTGCACAAATGCGAGGTGAAGAAATGAAGATTATTCCTGTAATAATTC 181834  
QY 121 ACAGAGCTTTAACAGCCCTGGAGAGCTCCCTGGGTGCTGTGAGACGACATGCTGAAGACA 180  
DB 181833 ACAGAGCTTTAACAGCCCTGGAGAGCTCCCTGGGTGCTGTGAGACGACATGCTGAAGACA 181774  
QY 181 TGATGCGTGTCTTCAAGACGAGTGTGTCAGAAAGTTGACCCATTGGAAACAGCAAG 240  
DB 181773 TGATGCGTGTCTTCAAGAAATGAGTTGTCAGAAAGTTGACCCATTGGAAACAGCAAG 181714  
QY 241 TGAATTTAGTTTCGCAATCACTTAATTCATGTTGGGTTATTTGGCACTCAAG 300  
DB 181713 TGAATTTAGTTTCGCAATCACTTAATTCATGTTGGGTTATTTGGCACTCAAG 181654  
QY 301 GAGTTATCCCAAGAGACATCCAGTGAAGACAGAACTGAAAGAAATCAGAGTCTACATGA 360  
DB 181653 GAGTTATCCCAAGAGACATCCAGTGAAGACAGAACTGAAAGAAATCAGAGTCTACATGA 181594  
QY 361 ACAGAGTTAAAGAAATTAACAGACAGAAAGAGCTGCCAAGCTGAGAGAGTCTGCTT 420  
DB 181593 ACAGAGTTAAAGAAATTAACAGACAGAAAGAGCTGCCAAGCTGAGAGAGTCTGCTT 181534  
QY 421 CGAGATTGTCAAGAGGACCTCTGGGAACCCAAACGAAAGACACACCAAAAGTGCCTA 480  
DB 181533 CGAGATTGTCAAGAGGACCTCTGGGAACCCAAACGAAAGACACACCAAAAGTGCCTA 181474  
QY 481 ATAAAGGAAAGCAACAC--TAATCTTTGGTTTGAATGATCATGTTTCAAAAAGT 537  
DB 181473 ATAAAGGAAAGCAACACCTAATTAATCTTTGGTTTGAATGATCATGTTTCAAAAAGT 181414  
QY 538 ACATCTCTTTTAATCAGTTTACATGATGATGATGATGATGATGATGATGATGATGATG 597  
DB 181413 ACATCTCTTTTAATCAGTTTACATGATGATGATGATGATGATGATGATGATGATGATG 181355  
QY 598 CTTTGGAAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 655  
DB 181354 CTTTGGAAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 181295  
QY 656 GGTGAGAAAGATTAAGTGTCTTTGTTGATTTTCAATTAAGATGATGATGATGATGATG 715  
DB 181294 GGTGAGAAAGATTAAGTGTCTTTGTTGATTTTCAATTAAGATGATGATGATGATGATG 181236  
QY 716 ATTGTAAGATTAATCTATAGCAGTGTGAAATCCAAATGTTCTCTGTAAACATTTGTAGT 775  
DB 181235 ATTGTAAGATTAATCTATAGCAGTGTGAAATCCAAATGTTCTCTGTAAACATTTGTAGT 181176  
QY 776 GTTGAAGAAATGAATATTTATGAAGTGTGATCTGTGATGATGATGATGATGATGATG 835  
DB 181175 GTTGAAGAAATGAATATTTATGAAGTGTGATCTGTGATGATGATGATGATGATGATG 181118  
QY 836 ATTTGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 894  
DB 181118 ATTTGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 181058

DB 181117 CATCTGTTTCAACATGATGAAAGTAAATGACTTAATTAATACCACATCTGTTCTGTTTC 181058  
QY 895 AGTTAGTCAACATGTTTCTGATATTTTCTTTTGTAGTAAATTTCTGTTGATATTC 954  
DB 181057 AGTTAGTCAACATGTTTCTGATATTTTCTTTTGTAGTAAATTTCTGTTGATATTC 181000  
QY 955 AAAGT-CAAAATGAAACCTTAAGGCTGTGACTTAATCTTCAATGTTCCATTTAAATTA 1013  
DB 180999 ACAGTACAAATTAATGAAACCTTAAGGCTGTGACTTAATCTTCAATGTTCCATTTAAATTA 180940  
QY 1014 AATGTTCTCATTAATCTGATGAAAT 1040  
DB 180939 AATGTTCTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 180913  
RESULT 7  
AC110304  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AC110304 224919 bp DNA linear HTG 15-NOV-2002  
Rattus norvegicus clone CH230-224B18, WORKING DRAFT SEQUENCE.  
AC110304.5 GI:25006992  
HTG; HTGS PHASE2; HTGS DRAFT; HTGS\_FULLTOP.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 224919)  
Mizny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alshrooke, S., Amin, A., Anguiano, D.,  
Avallebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, J., Cesar, H., Center, A.,  
Chaco, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Decker, D.,  
Delgado, O., Denson, S., Detamo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falter, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gedregotz, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,  
Gunnarsson, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Herrnstein, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
Hollins, B., Howell, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Joliver, A.,  
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C.L., Ledow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorenshew, L., Louised, H., Lozada, R.J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindaratne, M., Mahoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenan, B.,  
Milošević, A., Miner, G., Minya, S., Montemeyor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Muniz, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwankweli, O., Okunolu, G., Olarnunsgoon, A., Pal, S., Parks, K.,  
Pastermak, S., Paul, H., Perez, A., Perez, L., Frankoch, C.,  
Plopper, F., Polindexter, A., Popovic, D., Pritmus, E., Pol, L.,  
Pruzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
Sanders, W., Savary, G., Scherer, S., Scott, G., Shltsman, S., Shen, H.,  
Shetty, J., Shvartbeyn, A., Sison, I., Sitter, C.D., Smaj, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabot, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villaseña, D., Waldron, L., Walker, B., Wang, J.,  
Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willison, R., Wleczek, R., Wooden, H., Worley, K.,





Qy	398	CAACCTGGACAGAGAGCTGCTGCTGCAATTTGGTCAAGAGCACTCGGGAAACCAAGC	457
Db	24159	CAACCTGGACAGAGCTGCTGCTGCAATTTGGTCAAGAGCACTCGGGAAACCAAGC	2421
Qy	458	AAAAAGCAACCAAAAATGCGCTAATAAAGGAAAAAGCAACACTAATCTTTGGTTTGA	517
Db	24219	AAAAAGCAACCAAAAATGCGCTAATAAAGGAAAAAGCAACACTAATCTTTGGTTTGA	2427
Qy	518	TGTAACATGTTTCAAAAAGTACATCCTTTTAAATCAGTTTACAAATGATGTTATGTGACCA	577
Db	24279	TGTAACATGTTTCAAAAAGTACATCCTTTTAAATCAGTTTACAAATGATGTTATGTGACCA	2433
Qy	578	TGTGCTGTTTAAATGCAATTCCTTTTGGAAATTCATGATTAATTTTACCATTAACCTTTGTG	637
Db	24339	TGTGCTGTTTAAATGCAATTCCTTTTGGAAATTCATGATTAATTTTACCATTAACCTTTGTG	2439
Qy	638	ATACTGAATCTTTTGTGCTGAGAAAGATTAAAGTTGTCCTTTGTGATTTTCATATAAG	697
Db	24399	ATACTGAATCTTTTGTGCTGAGAAAGATTAAAGTTGTCCTTTGTGATTTTCATATAAG	2445
Qy	698	CATCATGATGTTTAAATTTATTTGTAAGATTTCTATAAGCAGTTGTGAAATCCAAATGTTG	757
Db	24459	CATCATGATGTTTAAATTTATTTGTAAGATTTCTATAAGCAGTTGTGAAATCCAAATGTTG	2451
Qy	758	TCTGTAAACATTTGTAGTGTTTGAATGAACCAATGATTTATGAGAGTGTCTATCTGTAG	817
Db	24519	TCTGTAAACATTTGTAGTGTTTGAATGAACCAATGATTTATGAGAGTGTCTATCTGTAG	2457
Qy	818	ACCTCGAGGGTGAAGCAATTTGTTTCACTAATGATGAGAAATACAGTACCTTAATAAC	877
Db	24579	A-CTGGAGGGTGAAGCAATTTGTTTCACTAATGATGAGAAATACAGTACCTTAATAAC	2463
Qy	878	CCAATCTG-TTTCGTGTCAGTTAGTTCACATGTTTCGTGATTTTTTTTTTTTGGAGT	936
Db	24638	CCAATCTGTTTTCGTGTCAGTTAGTTCACATGTTTCGTGATTTTTTTTTTTTGGAGT	2469
Qy	937	AATT-CTGTCTGATATTCAAAGT-CAAATTTGAACCTTAAGGCTGTACTTTAATCTT	994
Db	24698	AATTGCTGTCTGTATATTCAAAGTACAAATTTGAACCTTAAGGCTGTACTTTAATCTT	2475
Qy	995	CATGTTCCATTAAAAATAAATGTCCTCATTAATCTCGATGAGAAA	1040
Db	24758	CATGTTCCATTAAAAATAAATGTCCTCATTAATCTCGATGAGAAA	24803
RESULT 9			
AC044845/c			
LOCUS	233365 bp	DNA	linear HTG 23-MAR-2003
DEFINITION	Mus musculus chromosome 11 clone RP23-175Kx2 map 11, WORKING DRAFT		
SEQUENCE	7 unordered pieces.		
AC044845			
AC044845	4	GI:29164568	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 233365)		
TITLE	Mus musculus chromosome 11, clone RP23-175Kx2		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 233365)		
AUTHORS	Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,		
	Anderson, S., Baldwin, J., Barna, N., Baetien, V., Beda, F.,		
	Boguslavsky, L., Bouknight, B., Brown, A., Burkett, G.,		
	Campanaro, A., Casle, A., Choepel, Y., Colangelo, M., Collins, S.,		
	Collymore, A., Cooke, P., DeAngelis, K., Dewar, K., Diaz, J.S.,		
	Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,		
	Glaeser, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,		
	Grand-Pierre, N., Grant, G., Hages, B., Heatford, A., Horton, L.,		
	Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, J., Karatas, A.,		
	Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehnocky, J.,		
	Levine, R., Lieu, C., Liu, G., Locke, R., Macdonald, P., Markyus, N.,		

TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	3 (bases 1 to 235365)
AUTHORS	Birren, B., Nusbaum, C., Lander, E., Abouelella, A., Allen, N.,

TITLE	Direct Submission
JOURNAL	Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Mar 21, 2003 this sequence version replaced qi:1313929.

**COMMENT**

TITLE  
JOURNAL

McCarthy, M., McGowan, P., McGuck, A., McKernan, K., McPheters, R., Meldrum, J., Menues, L., Mihova, T., Miranda, C., Menga, V., Morrow, J., Murphy, T., Naylor, U., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olliaro, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmaz, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (12-APR -2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 235365)

Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N., Anderson, S., Arachchi, H.M., Barn, N., Bastien, V., Bloom, T., Bogunlavsky, L., Bonkshalter, B., Camarata, E., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeAtellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erikson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gada, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagiapan, D., Hacos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karacas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Mathews, C., McCarthy, M., Meldrum, J., Menues, L., Mihova, T., Menga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunghang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retts, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talmaz, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataratnam, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zempek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 23, 2003 this sequence version replaced gi:13123929.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center

```

Center Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L9172
Center clone name: 175 K 24
-----
Summary Statistics
Sequencing vector: M13; W7815, 32% of reads
Sequencing vector: Plasmid; n/a; 68% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960711
Consensus quality: 233275 bases at least Q40
Consensus quality: 233783 bases at least Q40
Consensus quality: 234252 bases at least Q20
Insert size: 220000; agarose-fp
Insert size: 234765; sum-of-contigs
Quality coverage: 13.9 in Q20 bases; agarose-fp
Quality coverage: 13.0 in Q20.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      9033: contig of 9033 bp in length
*      9034: gap of 100 bp
*      9133: gap of 100 bp
*      9134      32953: contig of 23820 bp in length
*      32954      33053: gap of 100 bp

```

FEATURES	
*	33054 87658: contig of 54605 bp in length
*	87659 87758: gap of 100 bp
*	87759 117237: contig of 23479 bp in length
*	117238 117337: gap of 100 bp
*	117338 155012: contig of 31675 bp in length
*	155013 155112: gap of 100 bp
*	155113 207954: contig of 58442 bp in length
*	207955 208054: gap of 100 bp
*	208055 233365: contig of 27311 bp in length
*	233366 Location/Qualifiers

Query Match	Best Local Similarity	61.7%; Score 642; DB 2; Length 235365;
Matches	692; Conservative	0; Mismatches 10; Indels 4; Gaps 4;
338	GGAAAGAAATCAGAGTCTACATGAACAGAGTTAAAGAAATACAGACAGAGAGGCTGC	397
211267	GGAAAGAAATCAGAGTCTACATGAACAGAGTTAAAGAAATACAGACAGAGAGGCTGC	211208
398	CAAGCTGGACAGAGTGTGCTTCAGAGTTGTCAAGAGGACATCTGGGAACCCAAAG	457
211207	CAAGCTGGACAGAGTGTGCTTCAGAGTTGTCAAGAGGACATCTGGGAACCCAAAG	211148
458	AAAAAGCACCAAAAGTGCGCTAATTAAGAGGAAAGCAAAACCTATCTTTGGTTTTGA	517
211147	AAAAAGCACCAAAAGTGCGCTAATTAAGAGGAAAGCAAAACCTATCTTTGGTTTTGA	211088
518	TGTACATGTTTTCAAAAAGTACATCTTTTAAATCAGTTTCAATGTAGTTATGTGACCA	577
211087	TGTACATGTTTTCAAAAAGTACATCTTTTAAATCAGTTTCAATGTAGTTATGTGACCA	211028
578	TGTGCTGTTTAAATGAGATTCCTTTTGAATTCATGATTAATTAACACTTACATTTGG	637
211027	TGTGCTGTTTAAATGAGATTCCTTTTGAATTCATGATTAATTAACACTTACATTTGG	210968
638	ATAGTGAATCTTTTTTGTGAGAGAAATTAAGTTGTCTTGTGATTTGATATATAAG	697
210967	ATAGTGAATCTTTTTTGTGAGAGAAATTAAGTTGTCTTGTGATTTGATATATAAG	210908
698	CATCATGATGTGTTAATATATGTAGAATATCTATAGACAGTTGGAATCCAAATGTTG	757
210907	CATCATGATGTGTTAATATATGTAGAATATCTATAGACAGTTGGAATCCAAATGTTG	210848
758	TCTGTAAACATTTGTAGTGTGTTGAATGAACATGATATTAATGAAGTGTCTATCTGAG	817
210847	TCTGTAAACATTTGTAGTGTGTTGAATGAACATGATATTAATGAAGTGTCTATCTGAG	210788



QY 818 ACCTGAGGTGTAGAGCAATTGTTTTCAGTAATGATGAGAATAATACAGTACTTAATAC 877  
DB 210787 A-CTGAGAGGTGTAGAGCAATTGTTTTCAGTAATGATGAGAATAATACAGTACTTAATAC 210729  
QY 878 CCACTGTG-TTCTGTTCAGTACTGATGACATGTTTGTGATTTTTTTTTTTTTTTAGT 936  
DB 210728 CCACTGTGTTTTCAGTACTGATGACATGTTTGTGATTTTTTTTTTTTTTTAGT 210669  
QY 937 AATT-CTGTCTGTATTCAGAGT-CAAAATTGAACCTTAAGGCTGACTTAATCTT 994  
DB 210668 AATTGCTGTCTGTATTCAGAGTCAAAATTGAACCTTAAGGCTGACTTAATCTT 210609  
QY 995 CATGTCATTTAAATTAATGTTCTCATTAATCTGATGAGAAA 1040  
DB 210608 CATGTCATTTAAATTAATGTTCTCATTAATCTGATGAGAAA 210563

RESULT 10  
AY302220 789 bp mRNA linear ROD 28-MAY-2003  
LOCUS AY302220  
DEFINITION Cricetus griseus nuclear DNA-binding protein CID mRNA, complete  
c8.  
ACCESSION AY302220  
VERSION AY302220.1 GI:31095451  
KEYWORDS Cricetus griseus (Chinese hamster)  
SOURCE Cricetus griseus  
ORGANISM Cricetus griseus (Chinese hamster)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetus.  
1 (bases 1 to 789)  
Guang, L., Masabumi, S. and Maru, Y.  
Differential display analysis of BCR-ABL-regulated genes  
Unpublished  
2 (bases 1 to 789)  
Guang, L., Masabumi, S. and Maru, Y.  
Direct Submission  
Submitted (18-MAY-2003) Dept. Genetics, The Institute of Medical  
Science, 4-6-1, Shiroyanada, Minato-ku, Tokyo 108-8639, Japan  
Location/Qualifiers  
1. 789  
/organism="Cricetus griseus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10029"  
/cell\_type="CHO"  
44..469  
/codon\_start=1  
/product="nuclear DNA-binding protein CID"  
/protein\_id="AAP43113.1"  
/db\_xref="GI:31095452"  
/translation="MAGGKNNEDYDVEIHSLALSSLGAVDDMLKTMWSVRNELL  
OKDPLEQAKVDIVASVTLNMFVYLATGVNPKKHPVQQLERIRVYNNRVKEITD  
KKKAAKLDRCGAASRPVNMALMEPKQKNTPRVANKSKSKH"

ORIGIN  
Query Match 54.3%; Score 564.8; DB 10; Length 789;  
Best Local Similarity 87.2%; Pired. No. 1.8e-102;  
Matches 680; Conservative 0; Mismatches 87; Indels 13; Gaps 5;

DB 182 TTGGACCACTGAGCAGAAAGAGTGGATTGTTCTTCATACATTGAATTCATG 241  
QY 276 TTTGGGTTTATTTGGCACTCAAGATTAAATCCAAAGACATCCAGTGAAGCAA 335  
DB 242 TTTGGGTTTATTTGGCACTCAAGATTAAATCCAAAGACATCCAGTGAAGCAA 301  
QY 336 CTGAAAGATTCAGAGTCTCATGAACAGAGTTAAAGAAATTAACAGCAAGAAAGGCT 395  
DB 302 TTGAAAGATTCAGAGTCTCATGAACAGAGTTAAAGAAATTAACAGCAAGAAAGGCT 361  
QY 396 GCCAAGCTGACAGAGTCTGCTTCAGATTGTCAAGAGCAGCTCGGAAACCCAA 455  
DB 362 GCCAAGCTGACAGAGCCTGCTTCAGATTGTCAAGATTGTCAAGAGCCTCGGAAACCCAA 421  
QY 456 CGAAAAAGCACACCAAAATGTCCTAAATAAAGGAAAAAGCAACATATCTTTGGTTTT 515  
DB 422 CAGAAAAACACCAAAATGTCCTAAATAAAGGAAAAAGTAAACATTAATCTTCGTTCT 481  
QY 516 GATGTACATGTTTCAAAAAGTACATCTTTTAAATCAGTT---TACATGTAATATGT 572  
DB 482 GATGTACATGTTTCAAAAAGTACATCTTTTAAATCAGTTTAAATTAAGTATGTTGTGT 541  
QY 573 GACCATGTGTTTAAATGATTTCTTTTGAATTCATGATTAATTAATTAACATTCAT 632  
DB 542 GGCATGCAAGTTTAAATGATTTCTTTTGAATTCATGATTAATTAATTAACATTCAT 601  
QY 633 TTG-----TGATGATGAATCTTTTGTGAGAAAGATTAAGTGTCTTTGTGAT 685  
DB 602 TTGATGATGATGAATCTTTTGTGAGAAAGATTAAGTGTCTTTGTGAT 661  
QY 686 TTT-CATATAACATCAT-GATGTGTTAAATTTGTAAGATTTCTTAAGCAGTTGTG 743  
DB 662 TTTCATATAACATCATGATGTTTAAATTTGTAAGATTTCTTAAGCAGTTGTG 721  
QY 744 AAATCCAAATGTT-CTGTGTAACATTTGTAAGTGTGAAATGAACATGATTTATGA 802  
DB 722 AAATCCAAATGTTCTTTGTAACATTTGTAAGTGTGAAATGAACATGATTTATGA 781

RESULT 11  
AC123163 209355 bp DNA linear HTG 13-MAY-2003  
LOCUS AC123163  
DEFINITION Rattus norvegicus clone CH230-16589, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 3 unordered pieces.  
ACCESSION AC123163  
VERSION AC123163.5 GI:30579265  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 209355)  
Muzny, D., Marie, M., Metzger, M., Lee, S., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Casar, H., Chen, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Bayes, K.,  
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregeorgis, E., Geer, K., Gill, R., Girdy, M., Guerra, W., Guevara, W.,  
Gubarene, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,



```

Qy      872 AAATACCACTCT-GTTTCGTTCAGTATGTTCAACATGTTTCGNGATTTTTCCTTTT 930
Db      193392 AAATACCACTCTGTTTCGTTCAGTATGTTCAACATGTTTC-TGGAATTTTTCCTTTT 193450
Qy      931 TTGAGTAATTCGTTCGTTCAGTATGTTCAAGT-CAAAATGAAACCTTAAGCTGATCTTAA 989
Db      193451 GGAAGTAATTCGTTCGTTCAGTATGTTCAAGTCAAAATGAAACCTTAAGCTGATCTTAA 193510
Qy      990 TTCTTCATGTTTCCTTTAAATTAATGTTTCATTAAT 1029
Db      193511 TTCTTCATGTTTCCTTTAAATTAATGTTTCATTAAT 193550

RESULT 12
AC097131 221055 bp DNA linear HTG 10-MAY-2003
LOCUS Rattus norvegicus clone CH230-49P11, *** SEQUENCING IN PROGRESS
DEFINITION *** 4 unordered pieces.
AC097131.7 GI:30520752
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 221055)
Muzny,D,Marle, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alsbrooks,S, Amin,A, Angiano,D,
Alyshebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryan,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A,
Chaco,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Devila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,
Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falle,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,P,
Gibbsgeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W,
Gunnarsson,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,J,
Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M,
Hollins,B, Howells,S, Hulik,S, Hume,J, Idlebird,D, Jackson,A,
Jackson,L, Jacob,L, Jiang,S, Johnson,B, Johnson,R, Jolivet,A,
Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,Y, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorenzuewa,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J,
Manjeshwar,M, Mahindartine,M, Mahmoud,M, Malloy,K, Mangum,A,
Mangum,B, Mapa,P, Martin,K, Martin,R, Martinez,E,
Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
Morgan,M, Morris,K, Morris,S, Muidasa,M, Murphy,M, Nair,L,
Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
Nwachileh,O, Okunolu,G, Olarnpusasoon,A, Pal,S, Parks,K,
Pasternak,S, Paul,H, Perez,A, Perez,L, Pfamkoch,C,
Plopper,F, Polndexter,A, Popovic,D, Primus,E, Pu,L-L,
Puzos,M, Quiroz,J, Rachin,B, Reeves,K, Regier,M,A, Reigh,R,
Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J,
Sanders,W, Savery,G, Scherer,S, Scott,G, Shatsman,S, Shen,H,
Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajs,D,
Sneid,A, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J,
Steinle,M, Strong,R, Sutton,A, Swatek,A, Tabor,P, Taylor,C,
Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Umanai,K,
Valas,R, Vera,V, Villalana,D, Waldron,L, Walker,B, Wang,J,
Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F,
Williams,G, Willison,R, Wleczky,R, Woodson,H, Morley,K,
Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V,

```

```

TITLE Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
JOURNAL Niederhausen,A., Weis,R., Smith,D.R., Holt,R.A., Smith,H.O.,
REFERENCE Weinstock,G. and Gibbs,R.A.
AUTHORS Direct Submission
2 (bases 1 to 221055)
TITLE Unpublished
JOURNAL 2 (bases 1 to 221055)
REFERENCE Submitted (11-OCT-2001) Human Genome Sequencing Center, Department
AUTHORS of Molecular and Human Genetics, Baylor College of Medicine, One
JOURNAL Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 221055)
REFERENCE Rat Genome Sequencing Consortium.
AUTHORS Direct Submission
JOURNAL Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:25091251.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GHXP
Center clone name: CH230-49P11

----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 197551 bases at least Q40
Consensus quality: 200898 bases at least Q30
Consensus quality: 202942 bases at least Q20
Estimated insert size: 205057; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 42645: contig of 42645 bp in length
* 42646 42745: gap of unknown length
* 42746 173337: contig of 130592 bp in length
* 173338 173437: gap of unknown length
* 173438 218852: contig of 45415 bp in length
* 218853 218952: gap of unknown length
* 218953 221055: contig of 2103 bp in length.
* Location/Qualifiers
  1. 221055
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-49P11"
    3702. 4811
    /note="wgs contig"
    51725. 52966
    /note="wgs_contig"

```

## ORIGIN

Query Match 44.9%; Score 467.2; DB 2; Length 221055;  
Best Local Similarity 87.7%; Pred. No. 2,7e-83;  
Matches 614; Conservative 0; Mismatches 73; Indels 13; Gaps 9;

QY 338 GGAAGAGATGAGTCTACATGAAACAGAGTTAAAGAAATPACAGACAGAGAAAGGCTGC 397  
DB 158779 GGAAGAGATCCGAGTCTACATGAAACAGAGTTAAAGAAATPACAGACAGAGAAAGGCTGC 158838

QY 398 CAAGCTGACAGAGGCTGCTTCCGAGATTGTCAAAGAGCACTCTGGAAACCCAAACG 457  
DB 158839 CAAGCTGACAGAGGCGCGCGCGAGATTGTCAAAGATGCACTCTGGAAACCCAAAC 158898

QY 458 AAAAAGACACCAAAAGTGGCTATTAAGGAAAGCAAAAC---TAATCTTTGGTTT 514  
DB 158899 AAAAAGACACCAAAAGTGGCTATTAAGGAAAGCAAAACCTAATTAATCTTTGGTTT 158958

QY 515 TGATGTACATGTTTCAAAAAGTACATCTTTTATCAGTTTACATGTAGTATGTGA 574  
DB 158959 TGATGTACATGTTTCAAAAAGTACATCTTTT-ATCAGTTTACAGTACTTATGTGA 159017

QY 575 CCATGGGTGTTTAAATGATTCCTTTGGAATTCATGTATTAATTTACACTTACA-TT 633  
DB 159018 CCATGGGTGTTTAAATGATTCCTTTGGAATTCATGTATTAATTTACACTTAAATTT 159077

QY 634 TGTGATCTGAATC--TTTTTTTGTGAGAAAGATTAAGTTGCTTGTGATTTCAT 691  
DB 159078 TGTGATCTGAATCTTTTTTTTTTGTGAGAAAGATTAAGTTGCTTGTGATTTCAT 159137

QY 692 ATTAAGCATCATGATGCTTTTAAATTTGTAGATATTCATATPAGCAGTTGGAATCCAA 751  
DB 159138 ATTAAGCATCATGATGCTTTTAAATTTGTAGATATTCATATPAGCAGTTGGAATCCAA 159196

QY 752 ATGTTCTCTGTAACATTTGTAGTGTGAAATGAACATGATATTAAGAGTGGCAT 811  
DB 159197 ATGTTCTCTGTAACATTTGTAGTGTGAAATG--TAGTGTATGATGAAGTGGCAT 159254

QY 812 CTGTGACCTCGAGGTGTAGAGCACTTTGTTTCAAGTATGAGAAATACAGTGAATT 871  
DB 159255 CTGTGACCTCGAGGTGTAGAGCACTGTTTTCACAAATGATGAGAAATGATGACTT 159314

QY 872 AAATACCACTCT-GTTCTGTGAGTGAAGTTCACAGTTCCTGATTTTCTTTT 930  
DB 159315 AAATACCACTCTGTTTCTGTGAGTTCACAGTTCCTGATTTTCTTTT 159373

QY 931 TTGAGTAATTTCTGTTGATATTCAGAGT-CAAAATTGAAACCTTAAGCTGACTTAA 989  
DB 159374 GGAGTAATTTCTGTTGATATTCAGAGTTCACAGTTCCTTAAGCTGACTTAA 159433

QY 990 TTCTTCATGTTCCATTTAAATTAATGTTCTCTTAAT 1029  
DB 159434 TTCTTCATGTTCCATTTAAATTAATTAATTTCTCTTAAT 159473

RESULT 13  
BD221634 1156 bp DNA linear PAT 17-JUL-2003  
LOCUS Definition Method of inducing apoptosis in cell.  
BD221634  
ACCESSION BD221634.1 GI:33031404  
VERSION UP 2002517192-A/1.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens (human)  
REFERENCE 1 (bases 1 to 1156)  
AUTHORS Rothbarth,K., Stammer,H., Werner,D. and Nalle,P.  
TITLE Method of inducing apoptosis in cell.  
JOURNAL Patient: JP 2002517192-A 1 18-JUN-2002;  
DEUTSCHES KREBSFORSCHUNGSZENTRUM STIFTUNG DES OFFENTLICHEN RECHTS,  
PETER NALLS  
COMMENT OS Homo sapiens (human)

PN JP 2002517192-A/1  
PD 18-JUN-2002  
PR 02-JUN-1999 JP 2000552267  
PR 03-JUN-1998 DE 198 24 811.3  
PI KÄRSTEN ROTHBARTH, HERMANN STAMMER, DIETER WERNER, PETER NALLS PC  
C12N15/09,A61K38/00,A61K48/00,A61P35/00,C12N15/00,A61K37/02 CC  
Method of inducing apoptosis in cell  
FH Key Location/Qualifiers  
FT CDS (118)..(540).  
FT mat\_peptide (118)..(540).  
FEATURES  
source  
1..1156  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 41.7%; Score 433.8; DB 6; Length 1156;  
Best Local Similarity 71.1%; Pred. No. 2,2e-76;  
Matches 662; Conservative 0; Mismatches 252; Indels 17; Gaps 6;

QY 66 GAGTGGTCAATATGCGAGGTGAGAAATGAATGAGATTATCCGTAGAAATTCACGAG 125  
DB 106 GAGCTGGCATATATGCGAGGTGAGAAATGAATGAGATTATCCGTAGAAATTCACGAG 165

QY 126 TCTTTAAGACCTCGAGAGCTCCCTGGTCTGTGTGAGCAGCATGCTGAAGCATGATG 185  
DB 166 TATTTGACAGCGTTTGAAGATTCATTTGCTGTGTGAGATGATGCTGAAGCATGATG 225

QY 186 GCTGTTTCTGAGAAACAGTGTGTCAGAGTTGAGACCCATGGAACAAGCAAGTGTGAT 245  
DB 226 TCTGTTTCTGAGAAATGATGTTGTCAGAGTTGAGATCCATGGAACAAGCAAGTGTGAT 285

QY 246 TTGTTTCTGATACCTTTAATTCATGTTTGGTTTATTTGGCACTCAAGAGATT 305  
DB 286 TTGTTTCTGATACCTTTAATTCATGTTTGGTTTATTTGGTTTATTTGGCAACCAAGAGATT 345

QY 306 AATCCCAAGAGATCCAGTGAAGAGAACTGGAAGAAATGATGATCTACATGAACGA 365  
DB 346 AATCCCAAGAGATCCAGTGAAGAGAACTGGAAGAAATGATGATCTACATGAACGA 405

QY 366 GTTAAAGAAATTAACAGACAGAAAGAGCTGCAAGCTGACAGAGGTGCTCTCGAGA 425  
DB 406 GTTAAAGAAATTAACAGACAGAAAGAGCTGCAAGCTGACAGAGGTGCTCTCGAGA 465

QY 426 TTGTTCAGAAAGCACTCTGGGAACCCAAAGCAAAAGCAACCAAAAGTGGTAAATAA 485  
DB 466 TTGTTCAGAAAGCACTCTGGGAACCCAAAGCAAAAGTGGTAAATAA 525

QY 486 GGAAGAAAGCAACATAATCTTTGTTGATGATGATGATGATGATGATGATGATGATGAT 544  
DB 526 GGAAGAAAGCAACATAATCTTTGTTGATGATGATGATGATGATGATGATGATGAT 585

QY 545 TTTTAATCAGTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 604  
DB 586 ATGTAATCAGTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 645

QY 605 AATCATGATATTAATTAACATTAATGATGATGATGATGATGATGATGATGATGATGAT 664  
DB 646 GAT-----TAGCACTGAGAGATTTTAACATGATGATGATGATGATGATGATGAT 700

QY 665 GATTAAGTTGCTTTGTTGATTTTCAATTAAGCATGATGATGATGATGATGATGATGAT 724  
DB 701 CTCTGATGAGACCTTAATTTCTTTATATAGGTGATGATGATGATGATGATGATGATGAT 760

QY 725 TATTCATTAACAGATTTGGAATCCAAATGTTCTCTGTAACATTTGATGTTGAAT 784  
DB 761 AGCTGTGAAA-----TTTAAGTAATGTTCTTTGTAACATTTGATGATTTTAAT 813

QY 785 GAACAAGATATTAATGAAGTGTGATCTATCTGATGATGATGATGATGATGATGATGAT 844  
DB 814 GAATTAATGATCTTATGAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 872

QY 845 CAGTAATGATGAGAAATACAGTACTTAATACCACCTCTCTTCTGTTAGTACTCA 904  
| | | | |  
DB 873 CACTA--TATGATATTAAGAAAGCGTGAATACCTTAATGCTATTTTCTGTATAGA 930  
| | | | |  
QY 905 ACATGTTTGGTATTTTTTTTTTTTGGAGTAAATCTGCTGATATTCAAAGT-CAA 963  
| | | | |  
DB 931 TACTTATCAATGTTTATGATGATTTAGGAATTAAGTCTGTTGTATATTCAAAGTGA 990  
| | | | |  
QY 964 ATGAACCTTAAGCGTACTTAATCTT 994  
| | | | |  
DB 991 ACTAAAGTTATGTTGTTACTTAATCTT 1021  
| | | | |  
RESULT 14  
AX009361 1156 bp DNA linear PAT 06-SEP-2000  
LOCUS AX009361  
DEFINITION Sequence 1 from Patent WO9963071.  
ACCESSION AX009361  
VERSION AX009361.1 GI:9996662  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1  
AUTHORS Rothbarth,K., Werner,D. and Stammer,H.  
TITLE Method for triggering apoptosis in cells  
JOURNAL Patent: WO 9963071-A 1 09-DEC-1999;  
DEUTSCHES KREBSFORSCH (DE); ROTHBARTH KARSTEN (DE); WERNER DIETER  
(DE); STAMMER HERMANN (DE)  
FEATURES  
source location/Qualifiers  
1..1156  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
118..543  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07450.1"  
/db\_xref="GI:9996663"  
/translation="MAGEEINEDYVPEIHLYLSAFENSIGAVDEMLTKMNSVSRNELL  
OKDPLEOAVDVSAATLNSMFVYATGKVPKEPVEKQELERIRVNNRVKEIRD  
KKKAAKGLDRGAARFVNALMPEKSNKASVANKSKS"  
ORIGIN  
Query Match 41.7%; Score 433.8; DB 6; Length 1156;  
Best Local Similarity 71.1%; Pred. No. 2.2e-76;  
Matches 662; Conservative 0; Mismatches 252; Indels 17; Gaps 6;  
QY 66 GAGTGTGCATGACGAGTGAAGAAATGAATTAATATCCCGTAGAATTCACGAG 125  
| | | | |  
DB 106 GACCTGCCATTAATGCGAGGTGAAGAAATTAATGAACATACATGACATGAATTCACGAG 165  
| | | | |  
QY 126 TCTTAAACAGCCCTGAGAGCTCCCTGGTGTGTGACGACATGCTGAAGACCATGATG 185  
| | | | |  
DB 166 TATTTCGACGCTTGAATTCATTTGCTGTGTGATGATGCTGAAGACCATGATG 225  
| | | | |  
QY 186 GCTGTTTCTGAAGACGATTTGTCAGAAAGTTGACCATTTGAACAAAGCAAGGTGAT 245  
| | | | |  
DB 226 TCTGTTTCTGAAGATGATTTGTCAGAAAGTTGATTCATTTGAACAAAGGTGAT 285  
| | | | |  
QY 246 TTGATTTTCTGATACATCTTAATTTCAATGTTTGGTTTATTTGGCAATCAAGAGTT 305  
| | | | |  
DB 286 TTGGTTTCTGATACATCTTAATTTCAATGTTTGGTTTATTTGGCAACCCAAAGAGTT 345  
| | | | |  
QY 306 AATCCCAAGAGCATCGAGTGAAGCAAGAACTGGAAGAAATCAAGATCTAATGAACAGA 365  
| | | | |  
DB 346 AATCTTAAGGAATCTCAGTAAACAGAAATTTGAAAGAAATTCAGAGTATTAATGAACAGA 405  
| | | | |  
QY 366 GTTAAAGAAATTAACAGCAAGAAAGAGCTGCCAAGCTGACAGAGGTGCTTTCCAGA 425  
| | | | |  
DB 406 GTCAAGAAATTAACAGCAAGAAAGAGCTGCCAAGCTGACAGAGGTGCTTTCCAGA 465  
| | | | |

QY 426 TTTCGAAGAGGACCTCTGGGAACCCAAAGAAAGACACCAAAAGTGGCTAATAAA 485  
| | | | |  
DB 466 TTTCGAAGAAATGCTCTGGGAACCCAAAGAAATTCGAAGAAATGATCAAAAGTTGCCAATAAA 525  
| | | | |  
QY 486 GGGAAAGCAACATATCTTTTGGTTTGAATGATCATGTTTCAAAAAGTACA-TCCT 544  
| | | | |  
DB 526 GGAAGAAAGTAAAGTTAACTTTTGGTTTGAATGATCAATATTCAAAAGTACATTAAT 585  
| | | | |  
QY 545 TTTTATCATGTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 604  
| | | | |  
DB 586 ATGTAATCAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 645  
| | | | |  
QY 605 AATCATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 664  
| | | | |  
DB 646 GAT-----TACACTGAGGATTTTACATTTGATTAATTAATTAATTAATTAATTAATTAAT 700  
| | | | |  
QY 665 GATTAAAGTGTCTTTTGTGATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 724  
| | | | |  
DB 701 CTCTGATGAGAGCTTAATTTCTTATATATAGTCAAGTCTTGAAGTCAATTTATATAGC 760  
| | | | |  
QY 725 TATTTCTAAGCAGTTGTGAATTCAAATGTTCTCTGTAACATTTGTAGTTGAAT 784  
| | | | |  
DB 761 AGCTGTGAAA-----TTTAAGTGAATGTTCTTTGTAACATTTGTACTAATTTAAAT 813  
| | | | |  
QY 785 GAACAATGATATTAATGAAGTGTCTGATCTGTAGACCTCGAGGTGAAGCAATTTGTTT 844  
| | | | |  
DB 814 GAATTAAGACCTTAATGAAGTGTCTGATCTGTAG- GCTGAATTAATTAATTAATTAATTAAT 872  
| | | | |  
QY 845 CAGTAATGATGAGAAATACAGTACTTAATACCACCTGTTTCTGTTAGTACTTCA 904  
| | | | |  
DB 873 CACTA--TATGATATTAAGAAAGCGTGAATGATCTTAATTTTCTGTATAGA 930  
| | | | |  
QY 905 ACATGTTTGGTATTTTTTTTTTTTGGAGTAAATCTGCTGATATTCAAAGT-CAA 963  
| | | | |  
DB 931 TACTTATCAATGTTTATGATGATTTAGGAATTAAGTCTGTTGTATATTCAAAGTGA 990  
| | | | |  
QY 964 ATGAACCTTAAGCGTACTTAATCTT 994  
| | | | |  
DB 991 ACTAAAGTTATGTTGTTACTTAATCTT 1021  
| | | | |  
RESULT 15  
AX329809 1172 bp DNA linear PAT 09-JAN-2002  
LOCUS AX329809  
DEFINITION Sequence 318 from Patent WO0194629.  
ACCESSION AX329809  
VERSION AX329809.1 GI:18102787  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1  
AUTHORS Young,P.B., Augustus,M., Carter,K.C., Ebner,R., Endreese,G.,  
Horrigan,S., Soppet,D.R. and Weaver,Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
JOURNAL Patent: WO 0194629-A 318 13-DEC-2001;  
Avalon Pharmaceuticals (US)  
FEATURES  
source location/Qualifiers  
1..1172  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 41.4%; Score 431; DB 6; Length 1172;  
Best Local Similarity 70.9%; Pred. No. 8e-76;  
Matches 660; Conservative 0; Mismatches 255; Indels 16; Gaps 6;  
QY 66 GAGTGTGCATATGCGAGGTGAAGAAATGAATGAAGATTAATCCGTAGAAATTCACGAG 125  
| | | | |  
DB 106 GACCTGCCATTAATGCGAGGTGAAGAAATTAATGAACATATTCAGTAAATTCACGAG 165  
| | | | |

OY	126	TCCTTAAACAGCCCTGGAAAGCTCCCTGGAGTGTGTGGACAGCATCTGAAAGCAATGATG	185
Db	166	TATTTGTCAAGCGTTTGAGAAATTCATTGGTGTGTGTGATGATGATGCTGAAGCCATGATG	225
OY	186	GCTGTTCCTAGAAACGAGTTGTGGAGAGTTGGACCCATTGGAAACAGCAAAAGTGAT	245
Db	226	TCTGTTTCTAAGAAATGAGTTGTGGAGAAAGTTGGATCCACTTGAACAGCAAAAGTGGAT	285
OY	246	TTAGTTCCTGATACACCTTAAATTCANGTGTTTGGGTTTATTTGGCAACTAGAGAGTT	305
Db	286	TTGGTTTCTGCAATACATTAATATCAATGTTTGGGTTTATTTGGCAACCAAGAGTT	345
OY	306	AATCCAAAGAGCATCCAGTAGAGCAGAACTGAAAGAAATCAGAGTCTACATGAAACAGA	365
Db	346	AATCCTAAGAAACATCCAGTAAACAGGAAATGGAAAGAAATCAGAGTATATATGAAACAGA	405
OY	366	GTTAAAGAAATPACAGACACAGAAAGGCTGCCAAGCTGGACACAGAGTGTCTCTTCGAA	425
Db	406	GTCAGGAAATPACAGACAGAAAGGCTGGCAAGCTGGACAGAGTGTCACTCTTAAGA	465
OY	426	TTTGTCAAGAAAGCACTCTGGGAAACCAACCAAAAGACACCAAAATGCTATTA	485
Db	466	TTTGTAAATAATGCCCCCTCTGGGAAACCAAAATGGAATAATGCAATCAAAATGTGCAATAA	525
OY	486	GCGAAAGCAAACTAATCTTTGGTTTGAATGATGATGTTTCAAAAAGTACA-TCCCT	544
Db	526	GGAATAAGTAAAGTTAACTTTTGGTTTGAATGATGATGATGATGATGATGATGATGAT	585
OY	545	TTTTAATCAGTTTACATGTAGTTATGTATGATGATGATGATGATGATGATGATGATGATG	604
Db	586	ATGTATATCAGATATATATATGTAAGCTAAATCTCTCTCCAAAGATCAATATCTTAT	645
OY	605	AATTCATGATTAATTTACATPACATTTGTGATATCTGAATCTTTTTTGTCTGAGAA	664
Db	646	GAT-----TAGACCTGAGGATTTTAACTGTGATATATATATATATATATATATACAT	700
OY	665	GATTAAAGTGTCTTTGTGATTTTCAATATAAAGCATGATGATGATGATGATGATGATGATG	724
Db	701	CTCTTGATGAGACTCTTATTTCTTTATATATAGGTCAAGTCTTGCAGATGACATTTATATAGC	760
OY	725	TATTCATATAGCAGTTGGGAAATCCAAATGTTCTCTGTAAACATTTGATGTTGAAAT	784
Db	761	AGCTGTGAAA-----TTTAAGTAAAGTTCTTTGTAAACATTTGTACTATATTTAAAT	813
OY	785	GAAACAATGATATATTAGAAGTGTGCTATCTGTAGACCTCGAGGCTGTAAAGACATTTGTTT	844
Db	814	GAAATATATGACTTAATGAAATATGCTATCTGTAG-GCTGAAATATATATGATACATCTGTTT	872
OY	845	CAGTATATGATGAATAATACAGTGACTTAAATACCACTCTGTTCTGTTCAAGTTAGTTCA	904
Db	873	CACATATATGATATTAAGAAAGCG-TGAATATGACTTAAATGTTTATTTTCTGTATAGA	931
OY	905	ACATGTTTCGTGATTTTTTTTTTTTTTGAATATTTCTGCTTGAATTCAAAGT-CAAA	963
Db	932	TACTTATATCATGTTTTCATGATTTTGAATATCTGCTTGTGATATTCAAAGTGTGAA	991
OY	964	ATTGAACCTTAAAGCTGTACTTAAATCTT 994	
Db	992	ACTAAAGTTTATGTGTGATCTTTAATTTCTT 1022	

Search completed: January 4, 2005, 11:20:39  
Job time : 4750.51 secs





CC cells resistant to conventional treatments. Overexpression of (I) is  
 CC sufficient itself to induce apoptosis but the effect may be increased  
 CC when used in combination with other anti-tumor methods. When cells  
 CC transfected with (I) undergo apoptosis, they release factors that kill  
 CC neighboring, non-transfected cells (bystander effect). This sequence  
 CC encodes the murine CID protein described in the method of the invention  
 XX  
 SQ Sequence 1038 BP; 323 A; 163 C; 219 G; 333 T; 0 U; 0 Other;

Query Match 99.8%; Score 1038; DB 3; Length 1038;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-230;  
 Matches 1038; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 3 GAACCCGTCATGCGTCATCATCGTGCACCATTTCCCGAGACAGCCGTCACGGT 62
Db 1 GAACCCGTCATGCGTCATCATCGTGCACCATTTCCCGAGACAGCCGTCACGGT 60
QY 63 ATTGAGTTGGTCACAAATGCGAGGTGAAGAAATGAAATGATATCCCGTAGAAATTCAC 122
Db 61 ATTGAGTTGGTCACAAATGCGAGGTGAAGAAATGAAATGATATCCCGTAGAAATTCAC 120
QY 123 GAGTCTTTAACAGCCCTGAGAGCTCCCTGGGCTGTGTGAGACATGCTGAAGACATG 182
Db 121 GAGTCTTTAACAGCCCTGAGAGCTCCCTGGGCTGTGTGAGACATGCTGAAGACATG 180
QY 183 ATGGCTGTTTCAAGAAACGAGTTGTGACAGAGTTGACCCCATTTGGAACAAGAGTGT 242
Db 181 ATGGCTGTTTCAAGAAACGAGTTGTGACAGAGTTGACCCCATTTGGAACAAGAGTGT 240
QY 243 GATTAGTTTCTGACATACCTTAATTTCAATGTTTGGGTTTATTTGGCACTCAAGCA 302
Db 241 GATTAGTTTCTGACATACCTTAATTTCAATGTTTGGGTTTATTTGGCACTCAAGCA 300
QY 303 GTTAATCCCAAGAGCATCAGTGAAGCAGAACTGGAAGAAAGATGAGTCTACATGAC 362
Db 301 GTTAATCCCAAGAGCATCAGTGAAGCAGAACTGGAAGAAAGATGAGTCTACATGAC 360
QY 363 AGAGTTAAAGAAATTAACAGAAAGAAAGGCTGCCAAGCTGAGCAGAGTGTCTCTCG 422
Db 361 AGAGTTAAAGAAATTAACAGAAAGAAAGGCTGCCAAGCTGAGCAGAGTGTCTCTCG 420
QY 423 AGATTGTCAAGAGCACTCTGGAGCCCAACGAAAGAACACACCCAAAGTGGCTAAT 482
Db 421 AGATTGTCAAGAGCACTCTGGAGCCCAACGAAAGAACACACCCAAAGTGGCTAAT 480
QY 483 AAAGGAAAGCAACACTAATCTTTGTTTGAATGACATGTTTCAAAAAGTACATC 542
Db 481 AAAGGAAAGCAACACTAATCTTTGTTTGAATGACATGTTTCAAAAAGTACATC 540
QY 543 CTTTAAATCAGTTTACATGTAATGTGACCATGCGGTAAATGATTCCTTTT 602
Db 541 CTTTAAATCAGTTTACATGTAATGTGACCATGCGGTAAATGATTCCTTTT 600
QY 603 GGAATTCATGTAATTAATTTACATTTGATGATCTGTAATCTTTTTCGTGAGA 662
Db 601 GGAATTCATGTAATTAATTTACATTTGATGATCTGTAATCTTTTTCGTGAGA 660
QY 663 AAGATTAGTGTCTTGTGATTTTCAATTAAGATCATGATGTTTATATTGTA 722
Db 661 AAGATTAGTGTCTTGTGATTTTCAATTAAGATCATGATGTTTATATTGTA 720
QY 723 GATTTCTTAAGACAGTTGGAATCCAAATGTTCTCTGTAACATTTGATGTTTGA 782
Db 721 GATTTCTTAAGACAGTTGGAATCCAAATGTTCTCTGTAACATTTGATGTTTGA 780
QY 783 ATGAACATGATTTTAAGAGTGTCTATCTGAGCTCTGAGGTGAAGACATTTGTT 842
Db 781 ATGAACATGATTTTAAGAGTGTCTATCTGAGCTCTGAGGTGAAGACATTTGTT 840
QY 843 TTCAAGTAATGATGAAGAAATACAGTGAATTAATACCACTCTGTTTCTGATTAAT 902
Db 841 TTCAAGTAATGATGAAGAAATACAGTGAATTAATTAATCACTCTGTTTCTGATTAAT 900

```

```

QY 903 CACATGTTTGTGTAATTTTTTTTTTTTGTAGTAATTCGTCTGTAAATTCAGAGTCAA 962
Db 901 CACATGTTTGTGTAATTTTTTTTTTTTGTAGTAATTCGTCTGTAAATTCAGAGTCAA 960
QY 963 AATTGAACCTTAAGGCTGTAATTTCTTCAATGTTCAATTAATTAATGTTTCTC 1022
Db 961 AATTGAACCTTAAGGCTGTAATTTCTTCAATGTTCAATTAATTAATGTTTCTC 1020
QY 1023 ATTAACTTGATGAGAAA 1040
Db 1021 ATTAACTTGATGAGAAA 1038

RESULT 2
ABL61981
ID ABL61981 standard; DNA, 1172 BP.
XX
AC ABL61981;
XX
DT 15-MAY-2002 (first entry)
XX
DE Colon adenocarcinoma related gene sequence SEQ ID NO:318.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;
KW cytosolic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN NO200194629-A2.
PD
13-DEC-2001.
XX
PF 30-MAY-2001; 2001WC-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.

```



PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 PA (AVALON PHARM.  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppel DR, Weaver Z;  
 DR WPI; 2002-188264/24.  
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.  
 XX  
 PS Claim 1; SEQ ID NO 318; 44bp; English.  
 XX  
 CC The present invention describes a method (M1) for screening for an anti-  
 CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in AB161664  
 CC to AB170110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms  
 CC tumour  
 CC  
 SQ Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;  
 Query Match 41.4%; Score 431; DB 6; Length 1172;  
 Best Local Similarity 70.9%; Pred. No. 1e-89;  
 Matches 660; Conservative 0; Mismatches 255; Indels 16; Gaps 6;  
 QY 66 GAGTTGTCATGAGGAGTGAAGAAATGAATGAATATATCCCGTGAATTCACAG 125  
 DB 106 GACCTGGCCATATGAGGAGTGAAGAAATGAATGAATATATCCAGTGAATTCACAG 165  
 QY 126 TCTTTAAGACCCCTGAGAGCTCCCTGGTGTGTGACGACATGCTGAAGACCATGATG 185  
 DB 166 TATTTCACACGCTTTGGAATTCATGCTGTGTGTGATGATGATGATGATGATG 225  
 QY 186 GCTGTTTCTGAAGAAGAGTGTGTGAGAGTGTGAGACCATGGAACAGCAAGAGTGAT 245  
 DB 226 TCTGTTTCTGAAGAATGAGTGTGTGAGAGTGTGAGACCATGGAACAGCAAGAGTGAT 285  
 QY 246 TTAGTTTCTGATACATCTTAATTAATTAATGTTTGGTTTATTTGGCACTCAAGAGTT 305  
 DB 286 TTGGTTTCTGATACATCAATTAATTAATTAATGTTTGGTTTATTTGGCAACCCAGAGATT 345  
 QY 306 AATCCCAAGAGCATCGAGTGAAGCAGAGCTGGAAGAAATCAAGTCTACATGAAGAGA 365  
 DB 346 AATCCTTAGAAGATCCAGTGAAGAAATGGAAGAAATCAAGTCTACATGAAGAGA 405  
 QY 366 GTTAAAGAAATTAAGAGAGAGAGTGTGCAAGTGTGAGAGTGTGCTTCGAGA 425  
 DB 406 GTCAAGAGAAATTAAGAGAGAGAGTGTGCAAGTGTGAGAGTGTGCTTCGAGA 465  
 QY 426 TTTGTCAAGAGAGAGTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485  
 DB 466 TTTGTCAAGAGAGAGTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 525  
 QY 486 GGGAAAGCAACATCAATCTTTGTTGATGATGATGTTTCAAAAGTACAC-TCCT 544

DB 526 GGAAGAAAGTAAAGTAACTTTTGGTTTGGATGATACATATTCAAAGATCAATTAAT 585  
 QY 545 TTTTAAACAGTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 604  
 DB 586 ATGTAATCAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 645  
 QY 605 AATTCATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 664  
 DB 646 GAT-----TAGCACTGAGATTTTAACTGATGATGATGATGATGATGATGATGAT 700  
 QY 665 GATTAAAGTGTCTTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 724  
 DB 701 CTCTGATGAGACCTTATTTCTTATATATGATGATGATGATGATGATGATGATGATGAT 760  
 QY 725 TATTCTAAGCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 784  
 DB 761 AGCTGTGAAA-----TTTAAAGTAAATGTTCTTGTAAACATTTGATGATTTTAAAT 813  
 QY 785 GAACATGATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 844  
 DB 814 GAATTAATGACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 872  
 QY 845 CAGTAATGATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 904  
 DB 873 CACTATGATGATTAAGAAAGCG-TGAAATGATGATTAATGATGATGATGATGATGATGAT 931  
 QY 905 ACATGTTGCTGATTTTTTTTTTTTTTTTGTGATGATGATGATGATGATGATGATGATGAT 963  
 DB 932 TACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 991  
 QY 964 ATTGAACCTTAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 994  
 DB 992 ACTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTT 1022  
 RESULT 3  
 ABK84717  
 ID ABK84717 standard; cDNA; 1172 BP.  
 AC ABK84717;  
 DT 14-AUG-2002 (first entry)  
 XX  
 DE Human cDNA differentially expressed in granulocytic cells #1288.  
 XX  
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
 KW viral infection; parasitic infection; protozoal infection;  
 KW fungal infection; sterile inflammatory disease; psoriasis;  
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KW adult respiratory distress syndrome; inflammatory bowel disease;  
 KW Crohn's disease; ulcerative colitis; periodontal disease;  
 KW granulocyte activation; chronic inflammation; allergy.  
 KW  
 OS Homo sapiens.  
 XX  
 PN W0200228999-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 03-OCT-2001; 2001WO-US030821.  
 XX  
 PR 03-OCT-2000; 2000US-0237189P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
 DR WPI; 2002-435328/46.  
 XX  
 PT Detecting granulocyte activation by detecting differential expression of  
 PT genes associated with granulocyte activation, which serves as diagnostic  
 PT markers that is useful for monitoring disease states and drug toxicity.

XX Claim 1; SEQ ID NO 1288; 114bp; English.

PS The invention relates to detecting (M1) granulocyte (GC) activation  
 XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing the  
 CC expression level to an expression level in an unactivated GC, where  
 CC differential expression of Gs is indicative of GCA. Also included are  
 CC modulating (M2) Gs by contacting GC with an agent that alters the  
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent  
 CC capable of modulating GCA or an inflammation (especially chronic) in a  
 CC tissue, an allergic response in a subject, exposure of a subject to a  
 CC pathogen or sterile inflammatory disease using the gene expression  
 CC profile; (3) detecting (M4) an inflammation (especially chronic) in a  
 CC tissue, an allergic response in a subject, exposure of a subject to a  
 CC pathogen or sterile inflammatory disease, by detecting a level of  
 CC expression in a sample of the tissue of gene(s) from Gs, where the level  
 CC of expression of the gene is indicative of inflammation; (4) treating  
 CC (M5) an inflammation (especially chronic) or in a tissue, an allergic  
 CC inflammatory disease, by exposing a subject to a pathogen or sterile  
 CC agent that modulates the expression of gene(s) from Gs in the tissue. M1  
 CC is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful  
 CC for screening an agent capable of modulating GCA preferably in an  
 CC (especially chronic) in a tissue; M4 is useful for detecting an inflammation  
 CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.,  
 CC perioritis, rheumatoid arthritis, renal reperfusion injury, ARDS, adult  
 CC respiratory distress syndrome, inflammatory bowel disease, Crohn's  
 CC disease, ulcerative colitis, peridontal disease, also bacterial  
 CC infection, viral infection, parasitic infection, protozoal infection,  
 CC fungal infection and M5 is useful for treating one of the above  
 CC conditions. The present sequence represents a gene differentially  
 CC expressed in granulocytes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;

XX Query Match 41.4%; Score 431; DB 6; Length 1172;  
 XX Best Local Similarity 70.9%; Pred. No. 1e-89;

XX Matches 660; Conservative 0; Mismatches 255; Indels 16; Gaps 6;

QY 66 GAGTTGTCACATGCGAGTGAAGAATGAATGATTTTCCTGAGAAATTCACGAG 125  
 DB 106 GAGTGGCCATATGCGAGTGAAGAATGAATGAATGATTTTCCTGAGAAATTCACGAG 165  
 QY 126 TCTTTAAAGCCCTGAGAGCTCCCTGGTGTCTGTGAGCAGCATGTGAAGACCAATGAG 185  
 DB 166 TATTGTGAGCGTTGAGAAATTCATGTGTCTGTGATGAGATGCTGAAGACCAATGAG 225  
 QY 186 GCTGTTTCTAGAAAGAGTTGTGAGAGTGAAGCCATTGGAACAGAAAGGTGAT 245  
 DB 226 TCTGTTTCTAGAAATGAGTTGTGAGAGTGAAGTGTGATCATTGGAACAGAAAGGTGAT 285  
 QY 246 TTAATTTCTGATACCTTAATTTCAATGTTTGGTTTATTTGGCACTCAAGAGTT 305  
 DB 286 TTGATTTCTGATACCTTAATTTCAATGTTTGGTTTATTTGGCACTCAAGAGTT 345  
 QY 306 AATCCAAAGAGATCCAGTGAAGCAGAACTGGAAGAAATGAGTGTACATGAACAGA 365  
 DB 346 AATCTTAAGAAATCCAGTGAAGCAGAACTGGAAGAAATGAGTGTACATGAACAGA 405  
 QY 366 GTTAAAGAAATTAACAGACAGAAAGGCTGCAAGCTGGAAGAGGTCTCTTGGAGA 425  
 DB 406 GTCAAGGAAATTAACAGACAGAAAGGCTGCAAGCTGGAAGAGGTCTCTTGGAGA 465  
 QY 426 TTGTCAAGAGGCACTTCTGGAAACCAAAAGCAACAAAGTGGCTATATAA 485  
 DB 466 TTGTCAAGAGGCACTTCTGGAAACCAAAAGTGGCTATATAA 525

QY 486 GCGAAAGCAACACTAATCTTTGGTTTGTGATGATGTTTCAAAAAGTACA-TCC 544  
 DB 526 GGAAGAAAGAAAGTAACTTTTGGTTTGTGATGATGATGATGATGATGATGATGAT 585  
 QY 545 TTTTAATGAGTTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 604  
 DB 586 AGTATATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 645  
 QY 605 AATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 664  
 DB 646 GAT-TTACAGTGAAGATTTTAACTGTGATGATGATGATGATGATGATGATGAT 700  
 QY 665 GATTAAAGTGTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 724  
 DB 701 CTCTGATGAGACTTATTTCTTATATGATGATGATGATGATGATGATGATGATGAT 760  
 QY 725 TATTCTTAAAGAGTGTGAAATCCAAATGTTCTGTGAAACATTTGTGATGATGAT 784  
 DB 761 AGCTGTGAAA-TTTAAGTGAAGTCTTTGTAAGATTTGTACTATTTTAAT 813  
 QY 785 GAACATGATATTAATGAGTGTCTATCTGTGAGCTCGAGGTGAAGACATTTGTTT 844  
 DB 814 GAATTAATGACCTTATGAGTGTATCTGTAG-GCTGAATTAATGATGATGATGATGAT 872  
 QY 845 CAGTAATGATGAGAAATACAGTGAATTAATCCGACTGTCTGTGATGATGATGATGAT 904  
 DB 873 CACTATATGATATTAAGAAAGG-TGAATGATTAATGATGATGATGATGATGATGAT 931  
 QY 905 ACATGTTTGTGATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 964  
 DB 932 TACTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 991  
 QY 964 ATTGAACCTTAAGCTGATCTTAATCTT 994  
 DB 992 ACTAAAGTTTATGTTGATCTTAATCTT 1022

RESULT 4

ID ADF81400 standard; DNA; 1172 BP.

XX ADF81400;

XX 26-FEB-2004 (first entry)

XX Leukaemia-related DNA sequence #1956.

XX Cytostatic; Gene therapy; leukaemia; ss.

XX unidentified.

XX WO2003039443-A2.

XX 15-MAY-2003.

XX 04-NOV-2002; 2002WO-EP012303.

XX 05-NOV-2001; 2001BP-00126244.

XX 30-APR-2002; 2002BP-00009758.

XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

XX (YULU-) UNIV LUDWIG MAXIMILIANS.

XX (HAFS/) HAFERLACH T.

XX (SCHO/) SCHOCH C.

XX (KERN/) KERN W.

XX Haferlach T, Schoch C, Kern W, Kohlmann A, Schittiger S, Dugas M;  
 PI Ellis R, Biers B, Mergenthaler S;  
 DR WPI; 2003-505037/47.

PT Determining the subtype of leukemia cells and whether a patient sample  
 contains leukemia cells or other cells, useful for treating leukemia,



Db 226 TCTGTTCTGGAATGAGTTGTCGAGAGTGGATCCACTGAAACAAGCAAAAGTGAT 285  
 Qy 246 TTAGTTCTGCATACACCTTAAATTCATGTTTGGGTTATTTGGCAACTCAAGAGATT 305  
 Db 286 TTGGTTCTGCATACACCTTAAATTCATGTTTGGGTTATTTGGCAACTCAAGAGATT 345  
 Qy 306 AATCCCAAGAGATCCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365  
 Db 346 AATCTTAAGAGATCCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405  
 Qy 366 GTTAAAGAAATTAACAGACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425  
 Db 406 GTCAAGAAATTAACAGACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465  
 Qy 426 TTTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485  
 Db 466 TTTGTAAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 525  
 Qy 486 GGGAAAGCAAGCACTAATCTTTGGTTTGTATGATGTTTCAAAAGTACA-TCC 544  
 Db 526 GAAAAAGTAAAGTTAACTTTGGTTTGTATGATGTTTCAAAAGTACAATTAAAT 585  
 Qy 545 TTTTATCAGTTTCAAGTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 604  
 Db 586 ATGTATACAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 645  
 Qy 605 AATTCAGTATTAATTTACACATTTACATTTGATGATGATGATGATGATGATGATGAT 664  
 Db 646 GAT-----TAGCATGAGAGATTTTAAACATGATGATGATGATGATGATGATGATGAT 700  
 Qy 665 GATTAGTTGCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 724  
 Db 701 CTCTGATGAGACTTATTTCTTTATTAAGTCACTGCTTGAAGTACATTTTATTAAGC 760  
 Qy 725 TATTCATACAGCTGTTGTAATCCAAATGTTCTCTGTAACATTTGATGATGATGATGAT 784  
 Db 761 AGCTGTAAG-----TTTAAAGTAAATGTTCTTTGTAACATTTGATGATGATGATGAT 813  
 Qy 785 GAAACAATGATATTAAGAGTGTATCTGTAGACTGAGGTGTAAGAGATTTGTTT 844  
 Db 814 GAATTAATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 872  
 Qy 845 CAGTATGATGAGAAATTAACAGTATTAATCCCACTGTTCTGTTGATGATGATGATGAT 904  
 Db 873 CACTATATGATATTAAGAAAGCG-TGAATGACTTAATGTTCAATTTTCTGATAGA 931  
 Qy 905 ACATGTTTCTGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 963  
 Db 932 TACTTATCATGTTTCTTCAATTTTGAATTTACTGCTTTGTTGATATTCAAAGTGA 991  
 Qy 964 ATTGAACCTTAAGGCTGTTACTTAAATCTT 994  
 Db 992 ACTAAAGTTTATGTTGTTACTTAAATCTT 1022

RESULT 6  
 AA243927 standard; cDNA; 1158 BP.  
 ID AA243927;  
 AC AA243927;  
 XX  
 DT 17-MAR-2000 (first entry)  
 XX Human C1D cDNA.  
 DE  
 XX C1D; human; apoptosis; tumour; gene therapy; treatment; ss.  
 OS Homo sapiens.  
 XX  
 FT Key Location/Qualifiers  
 CDS 118..543  
 FT /\*tag= a

FT /product= "C1D"  
 XX DE19824811-A1.  
 PN  
 XX  
 PD 09-DEC-1999.  
 XX  
 PF 03-JUN-1998; 98DE-01024811.  
 XX  
 PR 03-JUN-1998; 98DE-01024811.  
 XX  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
 PI Rothbarth K, Stemmer H, Werner D,  
 DR MPI; 2000-063506/06.  
 XX P-PSDB; AAY51024.  
 FT  
 FT Inducing apoptosis by overexpressing the C1D gene, particularly for  
 PS treating tumors.  
 XX  
 PS Claim 3; Fig 1; 10pp; German.  
 CC This invention describes a novel method for inducing apoptosis which  
 CC comprises overexpressing the C1D gene (I). The method is particularly  
 CC used to treat tumors and can also be used in gene therapy. The method has  
 CC no side effects on normal cells (contrast known methods of inducing  
 CC apoptosis such as cytotoxins and radiation), and may be effective on  
 CC cells resistant to conventional treatments. Overexpression of (I) is  
 CC sufficient itself to induce apoptosis but the effect may be increased  
 CC when used in combination with other anti-tumor methods. When cells  
 CC transfected with (I) undergo apoptosis, they release factors that kill  
 CC neighboring, non-transfected cells (bystander effect). This sequence  
 CC encodes the human C1D protein described in the method of the invention  
 SQ  
 Sequence 1158 BP; 388 A; 155 C; 237 G; 378 T; 0 U; 0 Other;  
 Query Match 41.3%; Score 430; DB 3; Length 1158;  
 Best Local Similarity 70.1%; Pred. No. 1.7e-89;  
 Matches 652; Conservative 0; Mismatches 265; Indels 13; Gaps 5;  
 Qy 66 GAGTTGCTCACAATGCGAGGAGAGAAATGAAATGAAATTAATCCGTAAGAAATTCACGAG 125  
 Db 106 GAGCTGGCCATTAATGAGAGAGAGAGAAATTAATGAAATTAATCCGTAAGAAATTCACGAG 165  
 Qy 126 TCTTAAAGCCCTGAGAGAGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 185  
 Db 166 TATTTGTCAGGCTTTGAGAAATTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 225  
 Qy 186 GCTGTTTCTGAAACGAGTGTGCAAGAGTTGGAACCCATGGAACAGCAAGGTGAT 245  
 Db 226 TCTGTTTCTGAAATAGTGTGTTGCAAGAGTTGATCCACTGGAACAGCAAGGTGAT 285  
 Qy 246 TTAGTTCTGCATACACCTTAAATTCATGTTTGGGTTATTTGGCAACTCAAGAGATT 305  
 Db 286 TTGGTTCTGCATACACCTTAAATTCATGTTTGGGTTATTTGGCAACTCAAGAGATT 345  
 Qy 306 AATCCCAAGAGATCCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365  
 Db 346 AATCTTAAGAGATCCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405  
 Qy 366 GTTAAAGAAATTAACAGACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425  
 Db 406 GTCAAGAAATTAACAGACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465  
 Qy 426 TTTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485  
 Db 466 TTTGTAAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 525  
 Qy 486 GGGAAAGCAAGCACTAATCTTTGGTTTGTATGATGTTTCAAAAGTACA-TTC 545  
 Db 526 GAAAAAGTAAAGTTAACTTTGGTTTGTATGATGTTTCAAAAGTACA---TT 582  
 Qy 546 TTTATCAGTTTCAAGTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 605

```

Db      583  AATATGTAATCAGAGTAATATGTAAGCTAAATCTTCCCAAGATCATTTATCTT 642
Qy      606  ATTGATGATTAATTAATTACATTAACATTTGATGATGAAATCTTTTTCGTAAGAAAG 665
Db      643  TATTGATTAAGACTGAGGATTTTAACATGATATATTAATTAATTAATTAACATC 702
Qy      666  ATTAAAGTGTCTTGTGATTTTTCATTAAGCATCATGATGATTTTAATTTGATGAT 725
Db      703  TCTTGATGAGACTCTTAATTTCTTAATTAAGGTGATGCTTGCAAGTACATTTTAAGCA 762
Qy      726  ATTCTAAGACAGAGTGTGAATCCAAATGCTCTGTAAGACATTTGATGATGAAATG 785
Db      763  GCGTGTGAA-----TTAAGTGAATGTTCTTTGTAACATTTGATGATTTTAATG 815
Qy      786  AACAAATATTAATTAAGATGATGATCTGATGACCTGAGGTGAAGACATTTGTTTC 845
Db      816  AATTAATGACCTTAATGAAGTATGATCTGATGAG-CTGAAATTAATGATACATCTGTTTC 874
Qy      846  AGTAATGATGAGAAATACAGTACTTAATTAATCCACATCTGTTCTGTCAGTTAGTCA 905
Db      875  ACTAATGATTAATTAAGAAAGCG-TGAATGACTTAATGTTCAATTTTCTGATAGAT 933
Qy      906  CATGTTTCGTGATTTTCTTTTCTTTTGTGAGTAATCTGCTGATTAATCAAGT-CAAAA 964
Db      934  ACTTATCATGTTTTCATGATTTTGAAGATTAATGCTGTTGTTGATTAATCAAGTGTGAAA 993
Qy      965  TTGAACCTTAAGGCTGATCTTAATTTCTT 994
Db      994  CTAAGAGTTATGTTGTTGATCTTAATTTCTT 1023

```

## RESULT 7

ADJ92841  
ID ADJ92841 standard; DNA; 426 BP.

XX ADJ92841:

XX 06-MAY-2004 (first entry)

XX Human co-repressor SUN-Cor DNA.

XX Nuclear receptor; therapy; diabetes; hyperlipidaemia; obesity;

XX atherosclerosis; human; co-repressor; gene; de.

XX Homo sapiens.

XX US2003228607-A1.

XX 11-DEC-2003.

XX 14-APR-2003; 2003US-00414692.

XX 15-APR-2002; 2002US-0372650P.

XX (WAGN/) WAGNER B L.

XX (SCHU/) SCHULMAN I G.

XX Wagner BL, Schulman IG;

XX WPI; 2004-167207/16.

XX Identifying compounds that bind to nuclear receptor and exhibit cell type

XX specific actions, and useful for treating hyperlipidemia, obesity and

XX diabetes.

XX Disclosure; SEQ ID NO 42; 99pp; English.

XX The invention relates to screening methods for identifying compounds that

XX bind to nuclear receptor and exhibit cell type specific actions. The

XX invention relates to modulators having an improved therapeutic profile.

XX The method is useful for identifying compounds that bind to a nuclear

XX receptor and exhibit cell type specific actions. It is also useful for

CC identifying modulators of nuclear receptors that are useful in treating

CC diseases e.g., diabetes, hyperlipidemia, obesity, atherosclerosis, etc.

CC The present sequence is human co-repressor DNA used to illustrate the

CC method of the invention.

CC Sequence 426 BP; 150 A; 76 C; 107 G; 93 T; 0 U; 0 Other;

Qy Query Match 40.2%; Score 418; DB 12; Length 426;

Qy Best Local Similarity 98.8%; Pred. No. 7.5e-87;

Qy Matches 421; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 78 ATGCGAGGTGAAGAAATGAATGAATATCCGTAGAAATTCACGAGCTTTTAACAGCC 137

Qy 1 ATGCGAGGTGAAGAAATGAATGAATATCCGTAGAAATTCACGAGCTTTTAACAGCC 60

Qy 138 CTGAGAGCTCCCTGGGTGCTGTGAGACGATGCTGAAGACATGATGCTGTTCTAGA 197

Qy 61 CTGAGAGCTCCCTGGGTGCTGTGAGATGATGATGCTGAAGACATGATGCTGTTCTAGA 120

Qy 198 AACGAGTTGTGCGAAGTTGGAACCCATGGAACAGCAAGGTGATTAAGTTTCTGCA 257

Qy 121 AATGAGTTGTGCGAAGTTGGAACCCATGGAACAGCAAGGTGATTAAGTTTCTGCA 180

Qy 258 TACACCTTAATTAATGATTTTGGGTTTATTTGGCAATCAAGGATTAATCCCAAGAG 317

Qy 181 TACACCTTAATTAATGATTTTGGGTTTATTTGGCAATCAAGGATTAATCCCAAGAG 240

Qy 318 CATTCAGTGAAGCAGAACTGGAAGAAATGAGATCTACATGAACAGATTAAAGAAATA 377

Qy 241 CATTCAGTGAAGCAGAACTGGAAGAAATGAGATCTACATGAACAGATTAAAGAAATA 300

Qy 378 ACAGACAAAGAAAGGCTGCCAAGCTGGAACAGAGGTGCTTGGATTTGCAAGAG 437

Qy 301 ACAGACAAAGAAAGGCTGCCAAGCTGGAACAGAGGTGCTTGGATTTGCAAGAG 360

Qy 438 GCACTCTGGAAACCAAAAGCAACCAAAAGCAACCAAAAGCAACCAAAAGCAACCAAAAG 497

Qy 361 GCACTCTGGAAACCAAAAGCAACCAAAAGCAACCAAAAGCAACCAAAAGCAACCAAAAG 420

Qy 498 CACTAA 503

Qy 421 CACTAA 426

Db

Db

Db

Qy

Qy

Qy

Qy

Qy

Qy

Qy

Qy

Qy

Qy

Qy

DR WPI; 2004-119264/12.  
XX  
PT New human genome-derived single exon nucleic acid probes useful for human  
PT gene expression analysis, for identifying or characterizing alternative  
PT splicing events, for assessing genomic alterations or as tools for  
PT surveying tissues.  
XX  
PS Claim 15; SEQ ID NO 1780; 80pp; English.  
XX  
CC The invention relates to a nucleic acid probe for measuring human gene  
CC expression, comprising any of the 27,400 fully defined nucleotide  
CC sequences in the specification, or their complements or fragments, and  
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences  
CC fully defined in the specification. The probe is a single exon probe that  
CC hybridizes under high stringency conditions to a nucleic acid molecule  
CC expressed in human cells or tissues. Also included are a spatially-  
CC addressable set of single exon nucleic acid probes for measuring human  
CC gene expression (comprising a plurality of single exon nucleic acid  
CC probes cited above), where each of the plurality of probes is separately  
CC and addressably isolatable or amplifiable from the plurality), a single  
CC exon microarray for measuring human gene expression, a method of  
CC measuring human gene expression, a vector comprising the single exon  
CC probe cited above, an ORF-encoded peptide comprising at least 8  
CC contiguous amino acids of any of the above-mentioned amino acid  
CC sequences (optionally with conservative amino acid substitutions), an  
CC isolated antibody that binds specifically to a peptide cited above,  
CC methods of selling and/or licensing single exon probes or microarrays to  
CC a customer desiring to measure gene expression, and a method of providing  
CC human gene expression data by subscription, and a computer-readable  
CC storage medium which contains a database having a plurality of records  
CC (each record including data on the expression of a single exon probe  
CC cited above. The probe, methods and apparatus are useful in gene  
CC expression analysis. The probes may be used as tools for surveying  
CC tissues to detect the presence of expressed messages that contain their  
CC specific exon, or in constructing genome-derived single exon microarrays.  
CC In addition, the probes are used in identifying and characterizing  
CC alternative splicing events, in detecting and characterizing gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe of the invention. Note: The present sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?docID=20030194704  
XX  
SQ Sequence 586 BP; 145 A; 138 C; 109 G; 194 T; 0 U; 0 Other;  
Query Match 37.8%; Score 392.6; DB 12; Length 586;  
Best Local Similarity 84.1%; Pred. No. 6.3e-81;  
Matches 443; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
XX  
QY 15 TGGGTCATCATCGTCGACCTATTTCCCGAGACAGGCGTCCAGGTTATGATTGTC 74  
DB |||||  
581 TGTGTCTATCTTCCGCCGACCGCTTTCCGGAGGCTGAGTCGAGGCCCTGAGTCAGCC 522  
QY 75 ACAATGGCAGGTGAAGAAATGAATGATGATATCCCGTAGAAATTCACGAGCTTTAAACA 114  
DB |||||  
521 ATAACGGCAGGTGAAGAAATTAATGAAGCATCTACATGAATTCACGAGTATTGTCA 462  
QY 135 GCCCTGAGAGCTCCCTGGGTGCTGTGACGACATGCTGAAGCAATGATGCTGTTTCT 194  
DB |||||  
461 GCGTTTAGAATTCATTGCTGTGTGATGAGATGCTGAAGCAACGATGCTGTTTCT 402  
QY 195 AGAAACGAGTGTGTCAGAGTGGACCCATTGGAACAAACGAAGTGAATTAAGTTTCT 254  
DB |||||  
401 AGAAATGAGTGTGTCAGAGTGGACCCATTGGAACAAACGAAGTGAATTAAGTTTCT 342  
QY 255 GCATACACCTTAATTCATGTTTGGTTTATTTGGCACTGACAGAGTTAATCCCAA 314  
DB |||||  
341 GCATACACCTTAATTCATGTTTGGTTTATTTGGCAACCAAGAGTTAATCTTAAG 282  
QY 315 GAGATCCAGTGAAGGAGCACTGGAAGAAATCAGAGTCTACATGAACAGAGTTAAAGAA 374  
|||

DB 281 GAACTCCAGTAAACAGAAATTGMAAGATCAGATATATATGAACAGATCAAGAA 222  
QY 375 ATAAACAGCAAGAAAGAGCTGCGCAGAGTGGATGCTGTTGAGATTGGTCAAG 434  
DB |||||  
221 ATAAACAGCAAGAAAGAGCTGCGCAGAGTGGATGCTGTTGAGATTGGTCAAG 162  
QY 435 AAGGCACTCTGGGAACCCAAACGAAAGACACACCAAAAGTGGCTAATTAAGGCAAGC 494  
DB |||||  
161 AATGCCCTCTGGGAACCCAAACGAAAGATGATCAAAAGTGGCAATTAAGGCAAGC 102  
QY 495 AAACACTAATCTTTGGTTTGGATGATCATGTTTCAAAAAGTACAT 541  
DB |||||  
101 AAAAGTTACTTTTGGTTTGGTTGTTGATCAATTTCAAAAAGTACAT 55  
RESULT 9  
AAL43974/C  
ID AAL43974 standard; cDNA; 426 BP.  
XX  
AC AAL43974;  
XX  
XX 27-SEP-2002 (first entry)  
XX  
DB Mammalian nuclear receptor cofactor CF7 reverse complement sequence.  
XX  
KW Mammalian; gene; 88; nuclear receptor cofactor; CF7; CF8;  
KW metabolism regulation; cell homeostasis; cell proliferation;  
KW differentiation; pathological cellular aberration;  
KW cellular defence mechanism.  
XX  
OS Mammalia.  
XX  
PN W0200242322-A2.  
XX  
PD 30-MAY-2002.  
XX  
PF 21-NOV-2001; 2001WO-EP013548.  
XX  
PR 21-NOV-2000; 2000EP-00125524.  
XX  
PA (LION-) LION BIOSCIENCE AG.  
XX  
PI Jackson D, Casari G, Suckow J;  
XX  
XX WPI; 2002-566559/60.  
DR  
XX  
PT Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators  
PT useful for inhibiting cellular function of cofactor and for treating  
PT metabolic disorders, immunological indications and hormonal dysfunctions.  
XX  
PS Claim 1; Fig 2; 68pp; English.  
XX  
CC The invention comprises the amino acid and coding sequences of two  
CC mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and  
CC CF8 protein sequences of the invention are useful for screening agents  
CC that are capable of inhibiting the cellular function of cofactor CF7  
CC and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes  
CC involved in cellular functions, such as: regulation of metabolism and  
CC cell homeostasis, cell proliferation and differentiation, pathological  
CC cellular aberrations, or cellular defence mechanisms. The present cDNA  
CC sequence represents the reverse complement of the mammalian nuclear  
CC receptor cofactor CF7 gene sequence  
XX  
SQ Sequence 426 BP; 105 A; 99 C; 64 G; 158 T; 0 U; 0 Other;  
Query Match 31.4%; Score 326.8; DB 6; Length 426;  
Best Local Similarity 85.4%; Pred. No. 9.5e-66;  
Matches 364; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
XX  
QY 78 ATGGCAGGTGAAGAAATGAATGATATCCCGTAGAAATTCACGAGCTTTAAACAGCC 137  
DB |||||  
426 ATGGCAGGTGAAGAAATTAATGAAGACTATTCATGAATTCACGATTATTGTCA 367  
|||





```
XX 22-SEP-2000; 2000EP-00120722.
PR (LION-) LION BIOSCIENCE AG.
PA
XX Casari G, Jackson D;
XX WPI; 2002-363179/41.
DR P-PSDB; ADI28083.
XX
PT Polynucleotide and polypeptide of novel nuclear receptor cofactor useful
PT for screening drugs regulating cofactor-associated physiological
PT responses e.g. hormonal dysfunctions.
XX
XX Claim 1; SEQ ID NO 1; 97bp; English.
XX
XX The invention relates to an isolated polynucleotide encoding a nuclear
XX receptor cofactor, also known as CF6. The polynucleotide or encoded
XX protein is useful for construction of multiple nuclear receptor cofactor
XX specific sequence alignments, preferably for the construction of protein
XX sequence alignments. The protein is useful for screening agents capable
XX of inhibiting the cellular function of the cofactor CF6. The
XX polynucleotide is useful for making vectors and for transforming cells,
XX both of which are ultimately useful for production of the CF6 protein.
XX They are also useful as scientific research tools for developing nucleic
XX acid probes for determining expression levels of the cofactor gene, e.g.
XX to identify diseased or otherwise abnormal states. They are particularly
XX useful for diagnostic purposes to e.g., identify deleted or mutant CF6
XX genes; or their measure expression. They are useful for developing
XX analytical tools such as antisense oligonucleotide for selectively
XX inhibiting expression of the cofactor gene to determine physiological
XX responses. The protein is useful for screening drugs for agonist and
XX antagonist activity, and therefore, for screening drugs for useful in
XX regulating physiological responses associated with the cofactors such as
XX neurosystemic diseases. The proteins are also useful for developing
XX antibodies for detection of the proteins. The polynucleotide can be used
XX to design primers for a polymerase chain reaction and are also used to
XX model the three-dimensional structure of the protein. This sequence
XX represents the sense strand of the CF6 gene coding region. (Note: this
XX sequence appears to be a fragment of the CF6 gene as the sequence differs
XX from the CF6 gene sequence given in Fig 1 (also designated SEQ ID NO: 1)
XX of the specification).
CC
CC Sequence 420 BP; 155 A; 64 C; 98 G; 103 T; 0 U; 0 Other;
SQ
Query Match 31.3%; Score 325.6; DB 6; Length 420;
Best Local Similarity 86.0%; Pred. No. 1.8e-65;
Matches 361; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
QY 78 ATGCGAGGTGAAAGAAATGAATGAATATCCCTAGAAATTCACGAGCTTTAAACAGCC 137
Db 1 ATGCGAGGTGAAAGAAATTAATGAAGACTATCCAAATTAATAACGATTTTGTGACGA 60
QY 138 CTGAGAGAGCTCCCTGGGTGCTGTGACGACATGCTGAAGACATGATGCTGTTCTAGA 197
Db 61 TTTCGGAATTCATTAATGATGCTGTGATGAGATGTGAAACATGATGCTGTTCTAGA 120
QY 198 AAGGAGTTGTCGAGAGTTGAGACCCATTGGAACAGCAAAAGGTGATTTAGTTTCTGCA 257
Db 121 AATGAGTTGTGCGAAGAGTTGAGACCCATTGGAACAGCAAAAGGTGATTTGTTCTGCA 180
QY 258 TACACCTTAATTAATCAATGTTTGGGTTTATTTGGCAACTCAAGAGTTAATCCCAAGNG 317
Db 181 TACACATTAATTAATCAATGTTTGGGTTTATTTGGCAACTCAAGAGTTAATCCCAAGNG 240
QY 318 CATCCAGTGAAGCAAGCAATCTGAAAGATCAAGTCTACATGAACAGAGTTAAAGAAATA 377
Db 241 CATCCAGTGAAGCAAGCAATCTGAAAGATCAAGTCTATATGAACAGAGTTAAAGAAATA 300
QY 378 ACAGACAAGAAAGAGCTGCCAAGCTGACAGAGGAGTCTTCAGATTTGTCAAGAAAG 437
Db 301 ACAGACAAGAAAGAGCTGCCAAGCTGACAGAGGAGTCTTCAGATTTGTAAAGAAAT 360
```

```
QY 438 GCACTCTGGGAACCCGAAACGAAAGCAACCAAAAGTGCATTAATGAAGGAAAGCAAA 497
Db 361 GCCCTCTGGGAACCAAAACGAAAGAAATATCAATTCCAAAAGTTGCCCAATGAAGAAAGTAA 420
RESULT 12
ADI28091/C
ID ADI28091 standard; DNA; 420 BP.
XX
XX ADI28091;
XX
XX 22-APR-2004 (first entry)
XX
XX Human nuclear receptor cofactor CF6 coding sequence complement strand.
XX
XX gene; nuclear receptor cofactor; CF6; cellular function inhibition;
XX metabolic disorder; immunological indication; hormonal dysfunction;
XX neurosystemic disease; ss.
XX
XX Homo sapiens.
XX
XX WO200224728-A2.
XX
XX 28-MAR-2002.
XX
XX 17-SEP-2001; 2001WO-EP010744.
XX
XX 22-SEP-2000; 2000EP-00120722.
XX
XX (LION-) LION BIOSCIENCE AG.
XX
XX Casari G, Jackson D;
XX WPI; 2002-363179/41.
XX
XX Polynucleotide and polypeptide of novel nuclear receptor cofactor useful
XX for screening drugs regulating cofactor-associated physiological
XX responses e.g. hormonal dysfunctions.
XX
XX Claim 1; SEQ ID NO 2; 97bp; English.
XX
XX The invention relates to an isolated polynucleotide encoding a nuclear
XX receptor cofactor, also known as CF6. The polynucleotide or encoded
XX protein is useful for construction of multiple nuclear receptor cofactor
XX specific sequence alignments, preferably for the construction of protein
XX sequence alignments. The protein is useful for screening agents capable
XX of inhibiting the cellular function of the cofactor CF6. The
XX polynucleotide is useful for making vectors and for transforming cells,
XX both of which are ultimately useful for production of the CF6 protein.
XX They are also useful as scientific research tools for developing nucleic
XX acid probes for determining expression levels of the cofactor gene, e.g.
XX to identify diseased or otherwise abnormal states. They are particularly
XX useful for diagnostic purposes to e.g., identify deleted or mutant CF6
XX genes; or their measure expression. They are useful for developing
XX analytical tools such as antisense oligonucleotide for selectively
XX inhibiting expression of the cofactor gene to determine physiological
XX responses. The protein is useful for screening drugs for agonist and
XX antagonist activity, and therefore, for screening drugs for useful in
XX regulating physiological responses associated with the cofactors such as
XX neurosystemic diseases. The proteins are also useful for developing
XX antibodies for detection of the proteins. The polynucleotide can be used
XX to design primers for a polymerase chain reaction and are also used to
XX model the three-dimensional structure of the protein. This sequence
XX represents the complement of the CF6 gene sense coding region.
XX
XX Sequence 420 BP; 103 A; 98 C; 64 G; 155 T; 0 U; 0 Other;
SQ
Query Match 31.3%; Score 325.6; DB 6; Length 420;
Best Local Similarity 86.0%; Pred. No. 1.8e-65;
Matches 361; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
```



QY 78 ATGGCAGTGAAGAAATGAAATGAAATATCCGTAGAAATTCAGAGTCTTTAAACAGCC 137  
 DB 420 ATGGCAGTGAAGAAATGAAATGAAATATCCGTAGAAATTCAGAGTCTTTAAACAGCC 361  
 QY 138 CTGGAGAGCTCCTGGTGTCTGTGAGCAGACATGCTGAAGACCATGATGCTGTTCTAGA 197  
 DB 360 TTGGCGAATTCATGATGCTGTGATGAGATGCTGAAGAACATGATGCTGTTCTAGA 301  
 QY 198 AAGAGTGTGTCAGAGTGTGAGACCCATGGAACAACGAAAGGTGATTTAGTTCTGCA 257  
 DB 300 AATGATGTGTGAGAGGTGAGACCCATGGAACAACGAAAGGTGATTTAGTTCTGCA 241  
 QY 258 TACACCTTAATCAATGATTTGGTATTATTTGCACTCAAGAGTAAATCCCAAGAG 317  
 DB 240 TACACCTTAATCAATGATTTGGTATTATTTGCACTCAAGAGTAAATCCCAAGAG 181  
 QY 318 CATTCAGTGAAGCAGAACTGGAAGAAATCAGAGTCTACATGAAACAGATTTAAAGAAATA 377  
 DB 180 CATTCAGTGAAGCAGAACTGGAAGAAATCAGAGTCTACATGAAACAGATTTAAAGAAATA 121  
 QY 378 ACAGACAAAGAAAGGCTGCCAAGCTGACAGAGGTGCTGTTGAGATTTGTCAAGAG 437  
 DB 120 ACAGACAAAGAAAGGCTGCCAAGCTGACAGAGGTGCTGTTGAGATTTGTCAAGAG 61  
 QY 438 GCACTGTGGGAAACCAACGAAAGAAAGCAACCAAGAGTGTATTAAGGAAAGCAA 497  
 DB 60 GCCCTCTGGGAAACCAACGAAAGAAAGTATCATCCAAAGTGTGCCATTAAGGAAAGTAAA 1

RESULT 13  
 AAL43975  
 ID AAL43975 standard; cDNA; 420 BP.

AC AAL43975;  
 DT 27-SEP-2002 (first entry)

DE Mammalian nuclear receptor cofactor CF8 coding sequence.

KM Mammalian; gene; 88; nuclear receptor cofactor; CF7; CF8;  
 KW metabolism regulation; cell homeostasis; cell proliferation;  
 KM differentiation; pathological cellular aberration;  
 KM cellular defence mechanism.

OS Mammalia.

Key Location/Qualifiers  
 CDS 1..420  
 FT /\*tag= a  
 FT /partial  
 FT /product= "Mammalian CF8 protein"  
 FT /note= "No start or stop codon is given"

PD WO200242322-A2.

PD 30-MAY-2002.

PF 21-NOV-2001; 2001WO-EP013548.

PR 21-NOV-2000; 2000EP-00125524.

PA (LION-) LION BIOSCIENCE AG.

PI Jackson D, Casari G, Suckow J;

DR WPI; 2002-566559/60.

DR P-PSDB; AAO15406.

PT Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators  
 PT useful for inhibiting cellular function of cofactor and for treating  
 PT metabolic disorders, immunological indications and hormonal dysfunctions.

PS Claim 1; Fig 1; 68bp; English.

XX The invention comprises the amino acid and coding sequences of two  
 CC mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and  
 CC CF8 protein sequences of the invention are useful for screening agents  
 CC that are capable of inhibiting the cellular function of cofactor CF7  
 CC and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes  
 CC involved in cellular functions, such as: regulation of metabolism and  
 CC cell homeostasis, cell proliferation and differentiation, pathological  
 CC cellular aberrations, or cellular defence mechanisms. The present cDNA  
 CC sequence encodes the mammalian nuclear receptor cofactor CF8 protein  
 XX

Sequence 420 BP; 153 A; 62 C; 100 G; 105 T; 0 U; 0 Other;

Query Match 30.3%; Score 314.6; DB 6; Length 420;  
 Best Local Similarity 84.7%; Pred. No. 6.3e-63;  
 Matches 353; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 81 GCAAGTGAAGAAATGAAATGAAATATCCGTAGAAATTCAGAGTCTTTAAACAGCCCTG 140  
 DB 1 GCAAGTGAAGAAATGAAATGAAATATCCGTAGAAATTCAGAGTCTTTAAACAGCCATTT 60  
 QY 141 GAGAGCTCCCTGGTGTCTGTGAGCAGACATGCTGAAGACATGATGCTGTTCTAGAAG 200  
 DB 61 GCGAATTCATGATGCTGTGATGAGATGCTGAAGAACATGATGCTGTTCTAGAAGAT 120  
 QY 201 GAGTTGTGAGAAAGTTGAGACCCATTGGAACAAGAAAGGTGATTTAGTTTCTGCATAC 260  
 DB 121 GAGTTGTGAGAAAGTTGAGACCCATTGGAACAAGAAAGGTGATTTAGTTTCTGCATAC 180  
 QY 261 ACCTTAATCAATGATTTGGTATTATTTGCACTCAAGAGTAAATCCCAAGAGCAT 320  
 DB 181 ATATTAAATTCATGATTTGGTATTATTTGCACTCAAGAGTAAATCCCAAGAGCAT 240  
 QY 321 CCAAGTGAAGCAGAACTGGAAGAAATCAGAGTCTACATGAAACAGATTTAAAGAAATACA 380  
 DB 241 TCAGTAAAGCAGAAATGGAAGAAATCAGAGTATATGAAACAGAGTCAAGAAATAGCA 300  
 QY 381 GACAAGAAAGAGCTGCCAAGCTGACAGAGGTGCTGTTGAGATTTGTCAAGAGGCA 440  
 DB 301 GACAAGAAAGAGCTGCCAAGCTGACAGAGGTGCTGTTGAGATTTGTCAAGAGGCA 360  
 QY 441 CTCTGGGAAACCAACGAAAGAAAGCAACCAAGAGTGTATTAAGGAAAGCAA 497  
 DB 361 CTCTGGGAAACCAACGAAAGAAATGCAATCCAAAGTGTCTTAAGGAAAGTAAA 417

RESULT 14  
 AAL43976/c  
 ID AAL43976 standard; cDNA; 420 BP.

AC AAL43976;

DT 27-SEP-2002 (first entry)

DE Mammalian nuclear receptor cofactor CF8 reverse complement sequence.

KM Mammalian; gene; 88; nuclear receptor cofactor; CF7; CF8;  
 KW metabolism regulation; cell homeostasis; cell proliferation;  
 KM differentiation; pathological cellular aberration;  
 KM cellular defence mechanism.

OS Mammalia.

PN WO200242322-A2.

PD 30-MAY-2002.

PF 21-NOV-2001; 2001WO-EP013548.

PR 21-NOV-2000; 2000EP-00125524.

PA (LION-) LION BIOSCIENCE AG.



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 4, 2005, 08:33:30 ; Search time 63.5 Seconds

(without alignments)  
1277.604 Million cell updates/sec

Title: US-09-701-618A-4  
Perfect score: 715  
Sequence: 1 MAGEENEDYPVEIHESLTA.....LMEPKRSTPKVANKGSKH 141

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_02:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	141	2	061368 mus musculu
2	704	98.5	141	2	035473 m small uni
3	684	95.7	141	2	07TSU0 cricetus
4	646	90.3	141	2	013901 homo sapien
5	515	72.0	111	2	07L8E1
6	182.5	25.5	141	2	086FR2 schistosoma
7	154.5	21.6	217	2	093VA9
8	144	20.1	204	2	06BUZ2
9	142	19.9	133	2	074469
10	137	19.2	207	2	06YX09
11	137	19.2	207	2	BAC9871 oryza sat
12	134	18.7	450	2	07SER2
13	131.5	18.4	169	2	07O1Z7
14	129	18.0	194	2	06C2H9
15	128	17.9	159	2	09VXL4
16	124	17.3	149	2	081L18
17	121.5	17.0	133	2	09N3J4
18	100	14.0	184	1	YH01 YEAST
19	98	13.7	149	2	06FUI1
20	97.5	13.6	654	2	08L2P4
21	90.5	12.7	244	2	09HGN2
22	90	12.6	395	2	06CQF7
23	89.5	12.5	364	2	086RM8
24	88	12.3	830	2	0872F6
25	87.5	12.2	177	2	06CMX3
26	87.5	12.2	229	2	08S0B1
27	87.5	12.2	479	2	06F182
28	86.5	12.2	479	2	09NQ22
29	86.5	12.1	508	1	GLPK MYCPN
30	86.5	12.1	595	2	086DH4
31	86.5	12.1	621	2	086RM2

32	86.5	12.1	625	2	086RM6	086rm6 plasmodium
33	86.5	12.1	639	2	086RM4	086rm4 plasmodium
34	86.5	12.1	732	2	086RM7	086rm7 plasmodium
35	86.5	12.1	849	2	086RM1	086r71 plasmodium
36	86.5	12.1	849	2	086RM3	086rm3 plasmodium
37	86.5	12.1	855	2	086DM2	086dm2 plasmodium
38	86.5	12.1	856	2	086RM6	086rm6 plasmodium
39	86.5	12.1	915	2	032863	032863 mycoplasma
40	86	12.0	1208	2	09VXU1	09vxu1 diosophila
41	86	12.0	1398	2	09VXU2	09vxu2 diosophila
42	85.5	12.0	328	2	0835W0	0835w0 enterococcu
43	85.5	12.0	370	2	06MR14	06mr14 bdellovibri
44	85.5	12.0	370	2	CAE77774	CAE77774 bdellovibri
45	85	11.9	1997	2	081AP3	081ap3 plasmodium

## ALIGNMENTS

RESULT 1	ID	PRELIMINARY	PRT	141 AA.
061368	061368			
AC	061368			
DT	01-NOV-1996 (TREMblrel. 01, Created)			
DT	01-NOV-1996 (TREMblrel. 01, Last sequence update)			
DT	01-MAR-2004 (TREMblrel. 26, Last annotation update)			
DE	CID protein.			
GN	Name=CID; Synonyms=CID;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NMRI; TISSUE=Ascites tumor;			
RX	MEDLINE=9816194; Pubmed=9469821;			
RA	Nehle P., Kecke T., Greferath R., Spiess E., Glaeser T., Rothbarth K.,			
RA	Stammer H., Werner D.,			
RT	"cDNA cloning, recombinant expression and characterization of			
RT	polypeptides with exceptional DNA affinity.";			
RL	Nucleic Acids Res. 26:1160-1166(1998).			
DR	EMBL; X95591; CAA64844.1; -			
DR	MGI; MGI:1927354; CID.			
DR	GO; GO:0017053; C:transcriptional repressor complex; IDA.			
DR	GO; GO:0003677; F:DNA binding; IDA.			
DR	GO; GO:0016922; F:lignand-dependent nuclear receptor interacto. . ; IDA.			
DR	GO; GO:0005515; F:protein binding; IPI.			
DR	GO; GO:0003714; F:transcription corepressor activity; IDA.			
DR	GO; GO:0016481; P:negative regulation of transcription; IDA.			
DR	InterPro; IPR011082; CID.			
DR	Pfam; PF07493; CID; 1.			
DR	SEQUENCE 141 AA; 16045 MW; C9AF68033A097161 CRC64;			
Query Match	100.0%; Score 715; DB 2; Length 141;			
Best Local Similarity	100.0%; Pred. No. 1.3e-55;			
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 MAGEENEDYPVEIHESLTALESSLGAVDMLKTMNAVSNNELLOKLDPLEQAKVDLSA 60			
DB	1 MAGEENEDYPVEIHESLTALESSLGAVDMLKTMNAVSNNELLOKLDPLEQAKVDLSA 60			
QY	61 YTNLSMFVYTLATQGNPKPKHPKQELERIRYVMNRVKEITDDKKAALKDRGAASRFVK 120			
DB	61 YTNLSMFVYTLATQGNPKPKHPKQELERIRYVMNRVKEITDDKKAALKDRGAASRFVK 120			
QY	121 ALMEPKRSTPKVANKGSKH 141			
DB	121 ALMEPKRSTPKVANKGSKH 141			
RESULT 2				
035473	PRELIMINARY; PRT; 141 AA.			

AC 035473;  
 DT 01-JAN-1998 (T-EMBLrel. 05, Created)  
 DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)  
 DT 01-OCT-2004 (T-EMBLrel. 28, Last annotation update)  
 DE Small unique nuclear receptor co-repressor (Nuclear DNA-binding  
 DE protein) (Mus musculus 10 days neonate skin cDNA, RIKEN full-length  
 DE enriched library, clone:4732439J08 product:nuclear DNA binding  
 DE protein, full insert sequence) (Mus musculus 12 days embryo embryonic  
 DE body between diaphragm region and neck cDNA, RIKEN full-length  
 DE enriched library, clone:9430096A21 product:nuclear DNA binding  
 DE protein, full insert sequence).  
 GN Name=Ctd; Synonyms=SN-COR;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OK NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98070763; PubMed=9405624;  
 RA Lazar M.A.;  
 RA Zamir I., Dawson J., Lavinsky R.M., Glass C.K., Rosenfeld M.G.,  
 RT "Cloning and characterization of a corepressor and potential component  
 RT of the nuclear hormone receptor repression complex."  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:14400-14405(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CZECH II;  
 RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datsenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.U., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McMan P.J., McKernan K.J., Malak J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.D., Lu X., Gibbs R.A.,  
 RA Fehy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Myers R.M., Butterfield Y.S.,  
 RA Rodriguez A.C., Skalska U., Schmutz J., Dickinson M.C.,  
 RA Krzywinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CZECH II;  
 RC TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;  
 RX STRAUSBERG R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning."  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=21085660; PubMed=11317851;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RA "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=20499374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes."  
 RL Genome Res. 10:1617-1630(2000).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanuki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:11757-1771(2000).  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanuki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:11757-1771(2000).  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanuki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:11757-1771(2000).  
 RN [11]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanuki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:11757-1771(2000).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanuki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:11757-1771(2000).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanuki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:11757-1771(2000).  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanuki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:11757-1771(2000).  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=C57BL/6J;  
 RC TISSUE=Embryonic body between diaphragm region and neck, and skin;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanuki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:11757-1771(2000).  
 RN [16]  
 RP SEQUENCE FROM N.A.

Query Match 98.5%; Score 704; DB 2; Length 141;  
 Best local Similarity 98.6%; Pred. No. 1.2e-54;  
 Matches 139; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MAGEENNEDYVPEIHESITALESSIGAVDDMLKTMMAVSRRELQKDPLEQAVDVLSA 60  
 DB 1 MAGEENNEDYVPEIHESITALESSIGAVDDMLKTMMAVSRRELQKDPLEQAVDVLSA 60  
 QY 61 YLTNSMFVYLTATGQVNVKHPVQKELRIRVNNRVKEITDKKAKALDRGAASRFVK 120

Db 61 YTLNSMFVYVLAATGQVNPKEHPVQELERIRVYNNRKYETIDKKKAAKLDGGAASRFVKN 120  
QY 121 ALWEPKRSKSTPKVANKGSKX 141  
Db 121 ALWEPKRSKSTPKVANKGSKX 141

RESULT 3  
Q7TSUO PRELIMINARY; PRT; 141 AA.  
ID Q7TSUO  
AC Q7TSUO  
DT 01-OCT-2003 (TReMBLrel. 25, Created)  
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE Nucleic acid-binding protein CID.  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetus.  
NCBI\_TaxID=10029;  
OK NCBI\_TaxID=10029;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Guang L., Masabumi S., Maru Y.;  
RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY302220; AAP4313.1; -;  
DR GO: 0003677; F:DNA binding; IEA.  
DR InterPro: IPR011082; CID.  
DR Pfam: PF07493; CID; 1.  
KM DNA-binding.  
SQ SEQUENCE 141 AA; 15945 MW; 0CF86B47764B6566 CRC64;

Query Match 95.7%; Score 684; DB 2; Length 141;  
Best Local Similarity 95.0%; Pred. No. 7.3e-53;  
Matches 134; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAGEBNMEDYPVETIHSLTALSSLGAVDDMLKTMAVSNRELLQKLDPEQAKVDLVA 60  
Db 1 MAGEBNMEDYPVETIHSLTALSSLGAVDDMLKTMAVSNRELLQKLDPEQAKVDLVA 60  
QY 61 YTLNSMFVYVLAATGQVNPKEHPVQELERIRVYNNRKYETIDKKKAAKLDGGAASRFVKN 120  
Db 61 YTLNSMFVYVLAATGQVNPKEHPVQELERIRVYNNRKYETIDKKKAAKLDGGAASRFVKN 120  
QY 121 ALWEPKRSKSTPKVANKGSKX 141  
Db 121 ALWEPKRSKSTPKVANKGSKX 141

RESULT 4  
ID Q13901 PRELIMINARY; PRT; 141 AA.  
AC Q13901  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-OCT-2004 (TReMBLrel. 28, Last annotation update)  
DE CID protein (Nucleic acid-binding protein).  
GN Name=CID;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
OK NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Term placenta;  
MEDLINE=98136194; Pubmed=9469821;  
RA Noble P., Kocks T., Greferath R., Spiess E., Glaeser T., Rothbarth K.,  
RA Stammer H., Werner D.;  
RT "cDNA cloning, recombinant expression and characterization of  
RT polypeptides with exceptional DNA affinity.";  
RL Nucleic Acids Res. 26:1160-1166(1998).  
RN [2]  
RP SEQUENCE FROM N.A.

RC TISSUE=Bone marrow, Lung, and Uterus;  
RX MEDLINE=22388257; Pubmed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huijk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A.C., Shcherchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skala U., Small D.E., Scherch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RA Strausberg R.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Strausberg R.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: X95592; CAA64845.1; -;  
DR EMBL: BC009584; AAH09584.1; -;  
DR EMBL: BC009589; AAH09589.1; -;  
DR EMBL: BC016284; AAH16284.1; -;  
DR EMBL: BC005235; AAH05235.1; -;  
DR GO: 0005634; C:nucleus; TAS.  
DR GO: 0003677; F:DNA binding; TAS.  
DR InterPro: IPR011082; CID.  
DR Pfam: PF07493; CID; 1.  
KM DNA-binding.  
SQ SEQUENCE 141 AA; 16019 MW; 9976A3BBD5620D63 CRC64;

Query Match 90.3%; Score 646; DB 2; Length 141;  
Best Local Similarity 90.0%; Pred. No. 1.7e-49;  
Matches 126; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 MAGEBNMEDYPVETIHSLTALSSLGAVDDMLKTMAVSNRELLQKLDPEQAKVDLVA 60  
Db 1 MAGEBNMEDYPVETIHSLTALSSLGAVDDMLKTMAVSNRELLQKLDPEQAKVDLVA 60  
QY 61 YTLNSMFVYVLAATGQVNPKEHPVQELERIRVYNNRKYETIDKKKAAKLDGGAASRFVKN 120  
Db 61 YTLNSMFVYVLAATGQVNPKEHPVQELERIRVYNNRKYETIDKKKAAKLDGGAASRFVKN 120  
QY 121 ALWEPKRSKSTPKVANKGSKX 140  
Db 121 ALWEPKRSKSTPKVANKGSKX 140

RESULT 5  
ID Q7L8E1 PRELIMINARY; PRT; 111 AA.  
AC Q7L8E1  
DT 05-JUL-2004 (TReMBLrel. 27, Created)  
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)

DE BA1501.1 (Novel protein similar to CID DNA-binding protein).  
 GN Name=BA1501.1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP NCBI\_TaxId=9606;  
 RP SEQUENCE FROM N.A.  
 RA Lawlor S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL354750; CAD13204.1;  
 DR GO; GO:0003677; P.DNA binding; IEA.  
 DR InterPro; IPR011082; CID.  
 DR Pfam; PF07493; CID; 1.  
 KW DNA-binding.  
 SQ SEQUENCE 111 AA; 12646 MW; F563EA631CECEE1 CRC64;

Query Match 72.0%; Score 515; DB 2; Length 111;  
 Best Local Similarity 92.7%; Pred. No. 5.6e-38;  
 Matches 102; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 31 MLKTMVAVSRLNLOKLDPLEQAKVDLSAYTNSMFWYLAQGVNPKKHPVKQELERI 90  
 DB 1 MLKTMVSRLNLOKLDPLEQAKVDLSAYTNSMFWYLAQGVNPKKHPVKQELERI 60  
 QY 91 RYVNRVKEITDKKKAKLDGASRTVKALMEPKKSTPKVANKGSK 140  
 DB 61 RYVNRVKEITDKKKAKLDGASRTVKALMEPKKSTPKVANKGSK 110

## RESULT 6

QY 086F82 PRELIMINARY; PRT; 141 AA.  
 ID 086F82;  
 AC 086F82;  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Clone Z2D1007 mRNA sequence.  
 OS Schistosoma japonicum (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 RN [1]  
 RP NCBI\_TaxId=6182;  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=22879925; PubMed=12973349;  
 RA Hu W., Yan Q., Shen D.K., Liu P., Zhu Z.D., Song H.D., Xu X.R.,  
 RA Wang Z.O., Kong Y.P., Zeng L.C., Wu J., Zhang X., Wang J.J., Xu X.N.,  
 RA Wang S.Y., Fu G., Zhang X.L., Wang Z.Q., Brindley P.J., McManus D.P.,  
 RA Xue C.L., Feng Z., Chen Z., Han Z.G.;  
 RT "Evolutionary and biomedical implications of a Schistosoma japonicum  
 RL complementary DNA resource."  
 RL Nat. Genet. 35:119-147(2003).  
 DR EMBL; AY222982; AAP06005.1;  
 DR InterPro; IPR011082; CID.  
 DR Pfam; PF07493; CID; 1.  
 SQ SEQUENCE 141 AA; 15966 MW; DBFPB20FPA1894BA0 CRC64;

Query Match 25.5%; Score 182.5; DB 2; Length 141;  
 Best Local Similarity 28.4%; Pred. No. 2.4e-08;  
 Matches 38; Conservative 33; Mismatches 60; Indels 3; Gaps 2;

QY 1 MAGEKNEDYVPEVTHESLITALESSLAGVDMLTMAVAARNELQKLDPLEQAKVDLSA 60  
 DB 1 MGSISIDFELIPKEISSGLVFSFATDVEQLVAKKISSFSNNS-SNEVSLDPTVKSELILC 59  
 QY 61 YLTNSMFWYLAQGVNPKKHPVKQELERIYVNRVKEITDKKKAK--LDGASRTVK 118  
 DB 60 YARNALPFMLLRONGVETQSHPTMDELDRVNNALKRCSGLVREVSALUTIDKKEITTFV 119  
 QY 119 KRALMEPKKSTPK 132  
 DB 120 KHALMKSATTKTK 133

RESULT 7  
 QY 093VA9 PRELIMINARY; PRT; 217 AA.  
 ID 093VA9;  
 AC 093VA9;  
 DT 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 05-UTL-2004 (TRENBLrel. 27, Last annotation update)  
 DE AT5G25080/711h3.90 (Hypothetical protein).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 RN [1]  
 RP NCBI\_TaxId=3702;  
 RP SEQUENCE FROM N.A.  
 RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Ban J.,  
 RA Bower L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Natusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=22088475; PubMed=12093376;  
 RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,  
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
 RT "Full-length messenger RNA sequences greatly improve genome  
 RL annotation."  
 RL Genome Biol. 3:RESEARCH0029-RESEARCH0029(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
 RA Feldmann K.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Shin P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,  
 RA Ban J., Bower L., Carninci P., Chung M.K., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Natusaka M.,  
 RA Nguyen M., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yu G., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY050469; AAK91482.1;  
 DR EMBL; AY087408; AAK64957.1;  
 DR EMBL; AF378870; AAK56573.1;  
 DR InterPro; IPR011082; CID.  
 DR Pfam; PF07493; CID; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 217 AA; 23748 MW; DB44F167BD773DE4 CRC64;

Query Match 21.6%; Score 154.5; DB 2; Length 217;  
 Best Local Similarity 30.6%; Pred. No. 1.2e-05;  
 Matches 41; Conservative 30; Mismatches 46; Indels 17; Gaps 3;

QY 22 ESSLGAVD-----MLKTMVAVSRLNLOKLDPLEQAKVDLSAYTNSMFWYLA 71  
 DB 18 ESIATENVNQTALYLKELRPOLEQMLTLAEFVLAAMPQLQAKTMHLAETITTVLELRL 77  
 QY 72 ATQGVNPKKHPVKQELERIYVNRVKEITDKKKAK--LDGASRTVK 126  
 DB 78 RCTGVDPDHRKKSIEIRINIVREKFCVQDSKGPLRPTTVLNQAAATRFIHSI--PD 135  
 QY 127 RSTPKVANKGSK 140  
 DB 136 LSTQKOSIRDLK 149

RESULT 8  
06BJ22 PRELIMINARY; PRT; 204 AA.  
AC 06BJ22;  
DT 01-OCT-2004 (TEMBLrel. 28, Created)  
DT 01-OCT-2004 (TEMBLrel. 28, Last sequence update)  
DE Similar to CA0533|IPF19872 Candida albicans IPF19872.  
GN ORFNames=DEHA0F27979;  
OS Debaryomyces hansenii (Yeast) (Torulasporea hansenii).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
OX NCBI\_TaxID=4959;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CB5767;  
RG GENOENVES;  
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,  
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
RA Barnay S., Blanchin S., Beckerich J.M., Beyne B., Bleykasten C.,  
RA Boistrasse A., Boyer J., Cactolico L., Confantollet F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Giropi A.,  
RA Hantaye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Niclaud J.M., Nikoleki M., Ozias S., Ozier-Kalogeropoulos O.,  
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,  
RA Swemene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
RA Zentou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,  
RA Bouchier C., Caudron B., Scarpell C., Gallardin C., Weissenbach J.,  
RA Mincker P., Soucier J.L.;  
RT "Genome evolution in yeasts";  
RL Nature 430:35-44(2004).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CB5767;  
RG Genoscope;  
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL, CR88238; CAG89900.1; -.  
SQ SEQUENCE 204 AA; 22569 MW; 724A5F9FC34D0C9C CRC64;

Query Match 20.4%; Score 144; DB 2; Length 204;  
Best Local Similarity 32.7%; Pred. No. 9.5e-05;  
Matches 35; Conservative 26; Mismatches 42; Indels 4; Gaps 2;

QY 20 ALESSGAVDDMLKTMVAASRNELQKLD-PLEQAKVDLSAYTLNMFVYIATGQVP 78  
DB 13 SLDSYDQLEKLEPIKSLSEKIASDSQVERIKYNNYSVYLISILPSYLTGINT 72

QY 79 KEHPVQELERIRVYNNRVETDK--KKAALDREGAARFVKAL 122  
DB 73 DQHPINKELTRIKLYMKRYKLEAKLASKOTSKEDAEAAFTIOMTL 119

RESULT 9  
074469 PRELIMINARY; PRT; 133 AA.  
AC 074469;  
DT 01-NOV-1998 (TEMBLrel. 08, Created)  
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)  
DE SPC173.07 protein.  
GN SPC173.07 protein.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972h-;  
RX MEDLINE=21848401; PubMed=11859360;  
Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.,  
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moulé S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S., Warren T., Whitehead S.,  
RA Woodward J., Volckarte G., Aert R., Robben J., Gymnopoulos B.,  
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Muller-Auer S.,  
RA Gabel C., Fuchs M., Dusterhoft A., Fritzc C., Holzer E., Meestl D.,  
RA Hilbert H., Borzym K., Langer I., Beck A., Lehman H., Reinhardt R.,  
RA Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta U.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Sipakovski G.V., Usery D., Barrett B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe";  
RL Nature 415:871-880(2002).  
RN [1]  
RP EMBL, AL031540; CAA20781.1; -.  
DR PIR; T41115; T41115.  
DR GenBank; Sponbe; SPC1739.07; -.  
DR InterPro; IPR011082; CID.  
DR Pfam; PF07493; CID; 1.  
SQ SEQUENCE 133 AA; 15068 MW; DD333501508D60F1 CRC64;

Query Match 19.9%; Score 142; DB 2; Length 133;  
Best Local Similarity 28.6%; Pred. No. 8.7e-05;  
Matches 38; Conservative 24; Mismatches 57; Indels 14; Gaps 3;

QY 13 EIHESLTALSSGAVDDMLKTMV-AVSRNELQKLDPLEQAKVDLSAYTLNMFVYI 71  
DB 4 EYSELPERLNKQDNVEDVLPKDBSIFELGSGSELOAKLYTWTSAINSTLSFY 63

QY 72 ATQGVNPKHPVQELERIRVYNNRV---KEITDKKAKAKLDREGAARFVKALMEPKR 127  
DB 64 KUNGIDASERPWNQELQRYGVNYISKIQAKKNNPTEAVNTSNAIS-----SSS 114

QY 128 KSTPKYANKGSK 140  
DB 115 SNRPKYAKDAATR 127

RESULT 10  
06YX09 PRELIMINARY; PRT; 207 AA.  
AC 06YX09;  
DT 05-JUL-2004 (TEMBLrel. 27, Created)  
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)  
DE Hypothetical protein OSUNBA0091D16.5.  
GN Name=OSUNBA0091D16.5;  
OS Oryza sativa (Japanese cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Eriactoidae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=3947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sasaki T., Marumoto T., Katayose Y.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AF005731; BAC99871.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 207 AA; 22203 MW; 0B0161AD8190E071 CRC64;

Query Match 19.2%; Score 137; DB 2; Length 207;  
Best Local Similarity 32.2%; Pred. No. 0.0004;



```

Matches 37; Conservative 24; Mismatches 44; Indels 10; Gaps 3

QY 18 LTALESSLGA---VDDMLKTMMAVSRN--ELLQKLDPLEQAKVDVLSAYTINSFWVYLA 72
Db 18 VSAADDTLAAABSVGDHIAEMVAAGAEPDIAIEPLPLRRARAFIAMAQAASLFAVRLR 77
QY 73 TGVNPKPEHPVQCELEIRIVYNNRVKELTDKKA-----AKIDRGAARFVKAL 122
Db 78 CSGVDPDDEHPKKEFERLSLMEEKLNRFEDMDKAPLPLPTTVVQTQAARFIQHSI 132

RESULT 11
BAC99871
ID BAC99871 PRELIMINARY; PRT; 207 AA.
AC BAC99871;
DT 02-MAR-2004 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein OSJNBA009J1D16.5.
GN OSJNBA009J1D16.5.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza; Oryza sativa.
CX NCBI_TaxID=39947;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
clone:OSJNBA009J1D16."
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005731; BAC99871.1; -.
SQ
Hypothetical protein.
KW
SEQUENCE 207 AA; 22203 MW; 0B0161AD9190E071 CRC64;

Query Match 19.2%; Score 137; DB 2; Length 207;
Best Local Similarity 32.2%; Pred. No. 0.0004;
Matches 37; Conservative 24; Mismatches 44; Indels 10; Gaps 3;

QY 18 LTALESSLGA---VDDMLKTMMAVSRN--ELLQKLDPLEQAKVDVLSAYTINSFWVYLA 72
Db 18 VSAADDTLAAABSVGDHIAEMVAAGAEPDIAIEPLPLRRARAFIAMAQAASLFAVRLR 77
QY 73 TGVNPKPEHPVQCELEIRIVYNNRVKELTDKKA-----AKIDRGAARFVKAL 122
Db 78 CSGVDPDDEHPKKEFERLSLMEEKLNRFEDMDKAPLPLPTTVVQTQAARFIQHSI 132

RESULT 12
Q7SER2
ID Q7SER2 PRELIMINARY; PRT; 450 AA.
AC Q7SER2;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=NCU02154.1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
NCBI_TaxID=5141;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Qui D., Iankovski P., Pedersen D., Nelson M., Washburne M.,
RA Seliterminkoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Roy A., Foley G., Mewes W., Staben C., Marcotte B., Greenberg D.,
RA Koy A., Foley K., Naylor U., Thomann N., Barrett R., Gierre S.,
RA Kamal M., Kamysvesels M., Maucelli E., Bieleke C., Rudd S., Frihman D.,

```

RA Kiyotofuoya S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,  
RA Croconi C., Marino G., Catchside D., Li W., Pratt R.J., Osmann S.A.,  
RA Desguza C.C., Glass L., Orzach M.J., Berglund J., Voelker R.,  
RA Varden O., Flammann M., Sæller S., Dunlap J., Radford A., Aramayo R.,  
RA Nativg D.O., Alex L.A., Mannheim P., Ebbole D.J., Freitag M.,  
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,  
RT "The Genome Sequence of the Filamentous Fungus *Neurospora crassa*." ;  
RL Nature 0:0-0(2003) .  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/genbank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC EMBL: AABX01000029; EAA35277.1; -.  
DR InterPro: IPR011082; CID.  
DR InterPro: IPR010625; CHCH.  
DR Pfam: PF07493; CID, 1.  
DR Pfam: PF06747; CHCH, 1.  
KM Hypothetical protein.  
SQ SEQUENCE 450 AA; 48201 MW; CQCF8F30E5AD0810 CRC64;

Query Match	18.7%;	Score 134;	DB 2;	Length 450;
Best Local Similarity	29.2%;	Pred. No. 0.0018;		
Matches 40;	Conservative 35;	Mismatches 54;	Indels 8;	Gaps 5;

Qy 9 DYPEIHESLALSSGAVDDMKTMVA5RNELLOQLDPLDEAKXNDL5A9YTLN5FW 68  
Db 2 DVP-DIRPLEKANVDLDQLEAALKPUG-DVGVSSKPLPLDKKLYMTVTAESTLIF 59  
Qy 69 VYLATQGVNPEHPEVKOLEKIRIRYNNRVKEITD-KKAAKLDGGA5RFVKALWEP 125  
Db 60 S5LNLNGDADEHAFELTETFRVROYFEKIONENPQOREQTVNKEAARFIRSDLDDE 119  
Qy 126 --KRKSTPKVANKGSK 140  
Db 120 AIRKQITELIA-KEKAK 135

RESULT 13	
070127	
ID 070127	PRELIMINARY; PRT; 169 AA.
AC 070127;	
DT 01-MAR-2004 (TrEMBLrel. 26, Created)	
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)	
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	
DE AGCP3387 (Fragment).	
GN Name=agcG52216; ORFNames=ENSGANG0000016976;	
OS Anopheles gambiae str. PE8T.	
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
CC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.	
OX NCBI_TaxID=180454;	
LN [1]	
RP	SEQUENCE FROM N.A.
RC STRAIN=PE8T;	
RA Anopheles Genome Sequencing Consortium;	
RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.	
-!- CAUTION: The sequence shown here is derived from an	
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is	
CC preliminary data.	
DR EMBL; AAA01008807; EAA04511.1; .-	
FT NON_TER	1
QO SEQUENCE	169 AA; 19590 MW; 69C2764F057A93C2 CRC64;

Query Match	18.4%;	Score 131.5;	DB 2;	Length 169;
Best Local Similarity	28.9%;	Pred. No. 0.00098;		
Matches	37;	Conservative	27;	Mismatches 55; Indels 9; Gaps 4;
QY	22	ESSIGAVDMLKTYMAVSRN-ELLQKLDPLEQAKVDLVSAATYNSMFVVYLANQG-VNPK	79	
	:	: :	: :	: :
	:	: :	: :	: :
	:	: :	: :	: :
Db	37	ENIMQAIDIQITLIANVRREYEKYPTFTBEKKRFPTLCAFCVNSLFWMHKEKRLGRNTV	96	
QY	80	EHPVKGELEIRIVVMNRVAEITDK-KKAALDRGAASRFPVKALWE-----PKRSSTPK	132	
	:	: :	: :	: :
	:	: :	: :	: :
	:	: :	: :	: :
Db	97	MDIKRKADLDVRVEAMKRQTIDHNLTKRRLLDQPAAGRVFAGLYIYANNAESRQKDAP	156	



```

Qy      133 VANKGSK 140
Db      157 NKKGRREK 164

RESULT 14
ID 06C2H9 PRELIMINARY; PRT: 194 AA.
AC 06C2H9;
DT 01-OCT-2004 (TREMBlrel. 28, Created)
DT 01-OCT-2004 (TREMBlrel. 28, Last sequence update)
DE Similar to DEHA0P271979 Debaryomyces hansenii.
GN ORFNames=YAL10P07559;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_Taxid=4952;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=CI8B9;
RG GENOLEVUES;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barney S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boistrume A., Boyer E., Cactolico L., Confalonieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Giropi A.,
RA Hantreay F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerret A., Kossul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Niclaud J.M., Nikolaki M., Oztas S., Ozler-Kalogeropoulos O.,
RA Pellens S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Suenne D., Tekala F., Wesolowski-Louvel M., Wescho E., Wirth B.,
RA Zentou-Meyer M., Zivanovic I., Holocin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
RA Wucher P., Soulet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=CI8B9;
RG Genoscope;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL: CR382132; CAG77940.1; D67D7D9F806EB3B CRC64;
SQ SEQUENCE 194 AA; 21520 MW; D67D7D9F806EB3B CRC64;

Query Match 18.0%; Score 129; DB 2; Length 194;
Best Local Similarity 28.3%; Pred. No. 0.0019;
Matches 32; Conservative 22; Mismatches 35; Indels 24; Gaps 3;

Qy      30 DMLKTMVAARN-----ELIQ-----KLDPLBQAKVDLVSAYTLNSFWVYLA 72
Db      3 DQIEDVLELSHNLQDVTSELSCQIKMIDFKGVAQLPDLBQAKVSKLAVYNSMFAFIL 62

Qy      73 TGVNPREHVKQELERIRVYNNRYKEITDK-----KKAALKDRGAARFV 118
Db      63 ASGADPTPIKMDLDRVKTVMGVAAHAGKPGPARDENTKVDPAARII 115

RESULT 15
ID 09VXL4 PRELIMINARY; PRT: 159 AA.
AC 09VXL4;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
GN ORFNames=CG8928;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;

```

```

RN
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; Pubmed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abtil J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Deitcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckner A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.G.,
RA Palazzolo M., Platan G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Regett K.C., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodagert, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao S., Zhu X., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; Pubmed=12537568;
RA Celniker S.E., Wheeler D.A., Kironmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; Pubmed=12537573;
RA Kironmiller J.S., Bergman C.M., Kironmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
a genomic perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; Pubmed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 4, 2005, 08:13:30 ; Search time 66 seconds  
(without alignments)  
766.376 Million cell updates/sec

Title: US-09-701-618a-4

Perfect score: 715  
Sequence: 1 MAGEENMEDPVEIHESLTA.....LMPEKRKSTPKVANKSKKH 141

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: A\_Geneseq\_23Sep04:\*  
2: geneseqp1980s:\*  
3: geneseqp1990s:\*  
4: geneseqp2000s:\*  
5: geneseqp2001s:\*  
6: geneseqp2002s:\*  
7: geneseqp2003as:\*  
8: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	715	100.0	141	3	AAV51025 Murine C1
2	704	98.5	141	7	ADBS9762 Rat Prote
3	646	90.3	141	3	AAV51024 Human C1D
4	646	90.3	141	7	ADBS9764 Human Pro
5	615	86.0	140	5	AD128083 Human nuc
6	615	86.0	141	5	AAO15405 Mammalian
7	579	81.0	140	5	AAO15406 Mammalian
8	485	67.8	102	3	AAO15406 Mammalian
9	235	32.9	54	8	ABO54730 Human sec
10	154.5	21.6	217	3	AAAG49051 Arabidops
11	154.5	21.6	217	3	AAAG49051 Arabidops
12	154.5	21.6	253	3	AAAG49050 Arabidops
13	154.5	21.6	256	3	AAAG49050 Arabidops
14	146.5	20.5	177	3	AAAG24126 Arabidops
15	146.5	20.5	177	3	AAAG24126 Arabidops
16	128	17.9	159	4	ABBB64082 Drosophi1
17	112.5	15.7	232	8	ABDP9133 C. albica
18	100	14.0	184	6	ABBS3602 Protein s
19	100	14.0	184	7	ADK64282 Disease t
20	87.5	12.2	479	3	AAV92242 Human can
21	87.5	12.2	479	5	ABBO5690 Human nuc
22	87.5	12.2	479	5	ABBB8498 Large hum
23	87.5	12.2	479	5	ABBP64721 Human pro
24	87.5	12.2	479	7	ADJ69180 Human hea
25	86.5	12.1	508	6	ABU36173 Protein e

26	86	12.0	1294	4	ABBB63502 Drosophi1
27	85.5	12.0	295	7	ADH85852 Enterococ
28	83.5	11.7	1055	2	AAW98865 H. pylori
29	82.5	11.5	2285	2	AAW98149 Bacillus
30	82	11.5	307	7	ADCC97413 E. faeciu
31	81	11.3	131	5	ABBP02625 Human ORF
32	81	11.3	345	3	AAAG23880 Arabidops
33	80.5	11.3	570	2	AAW57879 C. felis
34	80.5	11.3	570	2	AAW57860 C. felis
35	80.5	11.3	570	4	AAE12915 Ctenoceph
36	80.5	11.3	595	4	AAW57859 C. felis
37	80.5	11.3	595	2	AAW57861 C. felis
38	80.5	11.3	595	4	AAE12914 Ctenoceph
39	80.5	11.3	1094	5	ABP73717 Candida a
40	80.5	11.3	1976	7	ADBE63514 Rat Prote
41	80.5	11.3	1976	7	ADBE63518 Rat Prote
42	80	11.2	1027	2	AAAR43699 Protein L
43	80	11.2	1027	2	AAAR42203 Protein L
44	79	11.0	645	2	AAV35355 Amino aci
45	78.5	11.0	570	2	AAW57878 C. felis

ALIGNMENTS

RESULT 1	AAV51025 standard; protein; 141 AA.
XX	AAV51025;
AC	AAV51025;
XX	17-MAR-2000 (first entry)
DE	Murine CID protein.
XX	CID; murine; apoptosis; tumour; gene therapy; treatment.
XX	Mus sp.
OS	DE19824811-A1.
PN	09-DEC-1999.
PD	03-JUN-1998; 98DE-01024811.
XX	03-JUN-1998; 98DE-01024811.
PR	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA	Rothbarth K, Stammer H, Werner D;
XX	WPI; 2000-063506/06.
DR	N-PSDB; AA243928.
XX	Inducing apoptosis by overexpressing the CID gene, particularly for treating tumors.
PT	Claim 3; Fig 2; 10pp; German.
XX	This invention describes a novel method for inducing apoptosis which comprises overexpressing the CID gene (I). The method is particularly used to treat tumors and can also be used in gene therapy. The method has no side effects on normal cells (contrast known methods of inducing apoptosis such as cytotoxins and radiation), and may be effective on cells resistant to conventional treatments. Overexpression of (I) is sufficient itself to induce apoptosis but the effect may be increased when used in combination with other anti-tumor methods. When cells transfected with (I) undergo apoptosis, they release factors that kill neighboring, non-transfected cells (bystander effect). This sequence represents the murine CID protein described in the method of the invention
CC	Sequence 141 AA;
XX	
SO	

	Query Match	100.0%;	Score 715;	DB 3;	Length 141;
	Best Local Similarity	100.0%;	Pred. No. 4e-66;		
	Matches 141; Conservative	0;	Mismatches	0;	Indels
OY					Gaps
Dd	1 MAGEMNEDYPVEIHESLTALLESSLGAVIDDMLKTMVAVSERNELGKLPLEQAVYLISA	60			
OY	1 MAGEMNEDYPVEIHESLTALLESSLGAVIDDMLKTMVAVSERNELGKLPLEQAVYLISA	60			
Dd	61 YTLNSMRYVLYAQGVNPKHPVKOELERIRYYNRNVKEITDKKKAAATLDGAASRPVKK	120			
OY	61 YTLNSMRYVLYAQGVNPKHPVKOELERIRYYNRNVKEITDKKKAAATLDGAASRPVKK	120			
Dd	121 ALMEPRKRSTPKNVANKGSKH	141			
OY	121 ALMEPRKRSTPKNVANKGSKH	141			
Dd	121 ALMEPRKRSTPKNVANKGSKH	141			

  

RESULT 2

ADES9762 ID ADES9762 standard; protein; 141 AA.

XX AC ADE59762;

XX DT 29-JAN-2004 (first entry)

XX DE Rat Protein AAH05436, SEQ ID NO 5658.

XX KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;

XX KM chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX OS Rattus norvegicus.

XX PN W02003016475-AA2.

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002WC-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346383P.

XX PR 26-NOV-2001; 2001US-0333347P.

PA (GENO ) GEN HOSPITAL CORP.

PA (FARB ) BAYER AG.

XI Woolf C, D'urso D, Befort K, Costigan M;

PI MPI; 2003-268312/26.

DR GENBANK; AAH05436.

XX XX

New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a polynucleotide, a method for producing a pharmaceutical composition, a compound that regulates the activity of one or more of the polynucleotides, a method for identifying a compound or small molecule that regulates the activity of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating

CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 141 AA;

Query Match	98.5%	Score 704	DB 7	Length 141
Best Local Similarity	98.6%	Pred. No. 5,5e-65		
Matches 139	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY 1	MAGEBMNDYVEVETIHESITALESSIGAVDDMLKTMMAVSBNELLOKLDPLEQAKYDVLVA 60			
Db 1	MAGEBMNDYVEVETIHESITALESSIGAVDDMLKTMMAVSBNELLOKLDPLEQAKYDVLVA 60			
QY 61	YTLNSMFVWYLATQGVNPKEHPVQEOLEIRIRVYNNRVGEITDKKKAALDNRGAASRPVVK 120			
Db 61	YTLNSMFVWYLATQGVNPKEHPVQEOLEIRIRVYNNRVGEITDKKKAALDNRGAASRPVVK 120			
QY 121	ALMEPRKKSSTPKVANKKSKK 141			
Db 121	ALMEPRKKSSTPKVANKKSKK 141			

CC	RESULT 3
XX	AAYS1024
ID	AAYS1024 standard; protein; 141 AA.
AC	AAYS1024;
DT	17-MAR-2000 (first entry)
DE	Human C1D protein.
KW	C1D; human; apoptosis; tumour; gene therapy; treatment.
OS	Homo sapiens.
PN	DE19824811-A1.
PD	09-DEC-1999.
PJ	03-JUN-1998; 98DE-01024811.
PR	03-JUN-1998; 98DE-01024811.
PA	(DEKR-) DEUT KREBSFORSCHUNGSGEZENTRUM.
PI	Rothbarth K, Stammer H, Werner D;
DR	WPI; 2000-063506/06.
DR	N-PSDB; AAZ43927.
PT	Inducing apoptosis by overexpressing the C1D gene, particularly for
FT	treating tumors.
ES	Claim 3; Fig 1; 10pp; German.
XX	This invention describes a novel method for inducing apoptosis which
CC	comprises overexpressing the C1D gene (I). The method is particularly
CC	used to treat tumors and can also be used in gene therapy. The method has
CC	no side effects on normal cells (contrast known methods of inducing
CC	apoptosis such as cyclooxins and radiation), and may be effective on
CC	cells resistant to conventional treatments. Overexpression of (I) is
CC	sufficient itself to induce apoptosis but the effect may be increased
CC	when used in combination with other anti-tumor methods. When cells
CC	transfected with (I) undergo apoptosis, they release factors that kill

CC neighboring, non-transfected cells (bystander effect). This sequence  
CC represents the human CID protein described in the method of the invention  
XX  
SQ Sequence 141 AA;

Query Match 90.3%; Score 646; DB 3; Length 141;  
Best Local Similarity 90.0%; Pred. No. 5.9e-59;  
Matches 126; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 MAGEBNEDYVPEIHESLTALBESSLGAVDMLKTMAVSRNELQKLDPLEQAKVDLVA 60  
DB 1 MAGEBNEDYVPEIHESLTALBESSLGAVDMLKTMAVSRNELQKLDPLEQAKVDLVA 60  
QY 61 YTINSFWVYTLATQGVNPKKHPVKQELERIRVYNNRYKEITDDKKAKLDGGAASRFVKK 120  
DB 61 YTINSFWVYTLATQGVNPKKHPVKQELERIRVYNNRYKEITDDKKAKLDGGAASRFVKN 120  
QY 121 ALMEPKRKSTPKVANKGSKK 140  
DB 121 ALMEPKSKNSKVANKGSKK 140

RESULT 4  
ADES9764  
ID ADES9764 standard; protein; 141 AA.  
XX

AC ADES9764;  
XX  
XX 29-JAN-2004 (first entry)  
DT  
XX  
XX

DE Human Protein NP\_006324, SEQ ID NO 5660.

KM Human; pain; neuronal tissue; gene therapy;  
KM spinal segmental nerve injury; chronic constriction injury; CCI;  
KM spared nerve injury; SNI; Chung.

OS Homo sapiens.

PN WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002MO-US025765.

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

PA (GEHO) GEN HOSPITAL CORP.

PI (FARB) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M;

DR WPI; 2003-268312/26.

DR GENBANK; NP\_006324.

PT New composition comprising two or more isolated polypeptides, useful for  
XX preparing a medicament for treating pain in an animal.

PS Claim 1; Page; 1017pp; English.

CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a

CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

QY 1 MAGEBNEDYVPEIHESLTALBESSLGAVDMLKTMAVSRNELQKLDPLEQAKVDLVA 60  
DB 1 MAGEBNEDYVPEIHESLTALBESSLGAVDMLKTMAVSRNELQKLDPLEQAKVDLVA 60  
QY 61 YTINSFWVYTLATQGVNPKKHPVKQELERIRVYNNRYKEITDDKKAKLDGGAASRFVKK 120  
DB 61 YTINSFWVYTLATQGVNPKKHPVKQELERIRVYNNRYKEITDDKKAKLDGGAASRFVKN 120  
QY 121 ALMEPKRKSTPKVANKGSKK 140  
DB 121 ALMEPKSKNSKVANKGSKK 140

Query Match 90.3%; Score 646; DB 7; Length 141;  
Best Local Similarity 90.0%; Pred. No. 5.9e-59;  
Matches 126; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 MAGEBNEDYVPEIHESLTALBESSLGAVDMLKTMAVSRNELQKLDPLEQAKVDLVA 60  
DB 1 MAGEBNEDYVPEIHESLTALBESSLGAVDMLKTMAVSRNELQKLDPLEQAKVDLVA 60  
QY 61 YTINSFWVYTLATQGVNPKKHPVKQELERIRVYNNRYKEITDDKKAKLDGGAASRFVKK 120  
DB 61 YTINSFWVYTLATQGVNPKKHPVKQELERIRVYNNRYKEITDDKKAKLDGGAASRFVKN 120  
QY 121 ALMEPKRKSTPKVANKGSKK 140  
DB 121 ALMEPKSKNSKVANKGSKK 140

RESULT 5  
AD128083  
ID AD128083 standard; protein; 140 AA.  
XX

AC AD128083;

DT 22-APR-2004 (first entry)

DE Human nuclear receptor cofactor CF6 protein.

KM nuclear receptor cofactor; CF6; cellular function inhibition;

KM metabolic disorder; immunological indication; hormonal dysfunction;

XX neurosyretic disease.

OS Homo sapiens.

PN WO200224728-A2.

PD 28-MAR-2002.

PF 17-SEP-2001; 2001MO-EP010744.

PR 22-SEP-2000; 2000EP-00120722.

PA (LION-) LION BIOSCIENCE AG.

PI Casati G, Jackson D;

DR WPI; 2002-363179/41.

DR N-PSDB; AD128090.

PT Polynucleotide and polypeptide of novel nuclear receptor cofactor useful  
XX for screening drugs regulating cofactor-associated physiological  
XX responses e.g. hormonal dysfunctions.

PS Claim 12; SEQ ID NO 3; 97pp; English.  
XX The invention relates to an isolated polynucleotide encoding a nuclear

receptor cofactor, also known as CF6. The polynucleotide or encoded protein is useful for construction of multiple nuclear receptor cofactor specific sequence alignments, preferably for the construction of protein sequence alignments. The protein is useful for screening agents capable of inhibiting the cellular function of the cofactor CF6. The polynucleotide is useful for making vectors and for transforming cells, both of which are ultimately useful for production of the CF6 protein. They are also useful as scientific research tools for developing nucleic acid probes for determining expression levels of the cofactor gene, e.g. to identify diseased or otherwise abnormal states. They are particularly useful for diagnostic purposes to e.g., identify deleted or mutant CF6 genes; or their measure expression. They are useful for developing analytical tools such as antisense oligonucleotide for selectively inhibiting expression of the cofactor gene to determine physiological responses. The protein is useful for screening drugs for agonist and antagonist activity, and therefore, for screening for drugs useful in regulating physiological responses associated with the cofactors such as metabolic disorders, immunological indications, hormonal dysfunction, neurodegenerative diseases. The proteins are also useful for developing antibodies for detection of the proteins. The polynucleotide can be used to design primers for a polymerase chain reaction and are also used to model the three-dimensional structure of the protein. This sequence represents the CF6 protein. (Note: this sequence appears to be a fragment of the CF6 protein as the sequence differs from the CF6 protein sequence given in Fig 3 (also designated as SEQ ID NO: 2) of the specification).

Sequence 140 AA;

Query Match 86.0%; Score 615; DB 5; Length 140;  
Best Local Similarity 85.7%; Pred. No. 9.7e-56;  
Matches 120; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 1 MAGEENNEDYPVEIHESLTALSSIGAVDDMLKTMNAVSRNELQKDPLEQAVDVLSA 60  
DB 1 MAEEINEDYPVEIHDIYLSAFANSIDAVDEMKNMVSRYNELQKDPLEQAVDVLSA 60  
QY 61 YTLNSFWYLLATQGVNPEHPVKQELRIRVYNNRVEITDKKAAKLDGGAASRFVKK 120  
DB 61 YTLNSFWYLLATQGVNPEHPVKQELRIRVYNNRVEITDKKAAKLDGGAASRFVKN 120  
QY 121 ALMEPRKSTPKVANKGSK 140  
DB 121 ALMEPRKSTPKVANKGSK 140

RESULT 6  
AA015405  
ID AA015405 standard; protein; 141 AA.  
AC AA015405;  
DT 27-SEP-2002 (first entry)

Mammalian nuclear receptor cofactor CF7 protein.  
Mammalian nuclear receptor cofactor CF7 protein.

Mammalian; nuclear receptor cofactor; CF7; CF8; metabolism regulation;  
cell homeostasis; cell proliferation; differentiation;  
pathological cellular aberration; cellular defence mechanism.

Mammalia.

WO200242322-A2.  
30-MAY-2002.  
21-NOV-2001; 2001WO-EP013548.  
21-NOV-2000; 2000EP-00125524.  
(LION-) LION BIOSCIENCE AG.

Jackson D, Casari G, Suckow J;

DR WPI: 2002-566559/60.  
DR N-PSDB; AAL43973.  
XX  
PT Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators  
useful for inhibiting cellular function of cofactor and for treating  
PT metabolic disorders, immunological indications and hormonal dysfunctions.  
PS Claim 12; Fig 3; 68pp; English.

The invention comprises the amino acid and coding sequences of two mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and CF8 protein sequences of the invention are useful for screening agents that are capable of inhibiting the cellular function of cofactor CF7 and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes involved in cellular functions, such as: regulation of metabolism and cell homeostasis, cell proliferation and differentiation, pathological cellular aberrations, or cellular defence mechanisms. The present amino acid sequence represents the mammalian nuclear receptor cofactor CF7 protein.

Sequence 141 AA;

Query Match 86.0%; Score 615; DB 5; Length 141;  
Best Local Similarity 85.7%; Pred. No. 9.8e-56;  
Matches 120; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 1 MAGEENNEDYPVEIHESLTALSSIGAVDDMLKTMNAVSRNELQKDPLEQAVDVLSA 60  
DB 1 MAEEINEDYPVEIHDIYLSAFANSIDAVDEMKNMVSRYNELQKDPLEQAVDVLSA 60  
QY 61 YTLNSFWYLLATQGVNPEHPVKQELRIRVYNNRVEITDKKAAKLDGGAASRFVKK 120  
DB 61 YTLNSFWYLLATQGVNPEHPVKQELRIRVYNNRVEITDKKAAKLDGGAASRFVKN 120  
QY 121 ALMEPRKSTPKVANKGSK 140  
DB 121 ALMEPRKSTPKVANKGSK 140

RESULT 7  
AA015406  
ID AA015406 standard; protein; 140 AA.  
AC AA015406;  
DT 27-SEP-2002 (first entry)

Mammalian nuclear receptor cofactor CF8 protein.  
Mammalian nuclear receptor cofactor CF8 protein.

Mammalian; nuclear receptor cofactor; CF7; CF8; metabolism regulation;  
cell homeostasis; cell proliferation; differentiation;  
pathological cellular aberration; cellular defence mechanism.

Mammalia.

WO200242322-A2.  
30-MAY-2002.  
21-NOV-2001; 2001WO-EP013548.  
21-NOV-2000; 2000EP-00125524.  
(LION-) LION BIOSCIENCE AG.

Jackson D, Casari G, Suckow J;  
WPI: 2002-566559/60.  
DR N-PSDB; AAL43975.

Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators  
useful for inhibiting cellular function of cofactor and for treating  
PT metabolic disorders, immunological indications and hormonal dysfunctions.

XX  
PS Claim 12; Fig 3; 68bp; English.  
XX  
CC The invention comprises the amino acid and coding sequences of two  
CC mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and  
CC CF8 protein sequences of the invention are useful for screening agents  
CC that are capable of inhibiting the cellular function of cofactor CF7  
CC and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes  
CC involved in cellular functions, such as: regulation of metabolism and  
CC cell homeostasis, cell proliferation and differentiation, pathological  
CC cellular aberrations, or cellular defence mechanisms. The present amino  
CC acid sequence represents the mammalian nuclear receptor cofactor CF8  
CC protein  
XX  
SQ Sequence 140 AA;  
  
Query Match 81.0%; Score 579; DB 5; Length 140;  
Best Local Similarity 82.0%; Pred. No. 5,4e-52;  
Matches 114; Conservative 9; Mismatches 16; Indels 0; Gaps 0;  
  
QY 2 AGSENNEDYVEIHESITLAESSLGAVDDMLKTMNAVSRNELLQKLDPLEQAKVDLSAY 61  
DB 1 AAEINEDYVEIHEDYLSAFANSIDAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLSAY 60  
QY 62 TLNSMFVVYLTATGQVNPKEHPVKOEIRIRVYNNRVKEITDKKKQAKLDGGAASRPVYKA 121  
DB 61 TLNSMFVVYLTATGQVNPKEHPVKOEIRIRVYNNRVKEITDKKKQAKLDGGAASRPVYKA 120  
QY 122 LMPPKRSKTPKVNKSKK 140  
DB 121 LMPPKRNKSKVAVYKSKK 139  
  
RESULT 8  
AAG01231  
ID AAG01231 standard; protein; 102 AA.  
XX  
AC AAG01231;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein, SEQ ID NO: 5312.  
XX  
KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping.  
XX  
OS Homo sapiens.  
XX  
PN EP1033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-00200610.  
XX  
PR 26-FEB-1999; 99US-0122487P.  
XX  
PA (GEST) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
DR WPI; 2000-500381/45.  
XX N-PsDB; AAC01237.  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX  
PS Claim 13; SEQ ID NO 5312; 71bp + Sequence Listing; English.  
XX  
CC The present sequence is a polypeptide encoded by one of a large number of  
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different  
CC tissues. EST sequences usually correspond mainly to the 3' untranslated

CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
CC sequences derived from the 5' ends of mRNAs and even in those cases where  
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
CC are also used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. They are used to obtain upstream regulatory sequences  
CC and to design expression and secretion vectors  
XX  
SQ Sequence 102 AA;  
  
Query Match 67.8%; Score 485; DB 3; Length 102;  
Best Local Similarity 92.2%; Pred. No. 2.1e-42;  
Matches 94; Conservative 6; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 MAGEENNEDYVEIHESITLAESSLGAVDDMLKTMNAVSRNELLQKLDPLEQAKVDLSA 60  
DB 1 MAGEINEDYVEIHESITLAESSLGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLSA 60  
QY 61 YTLNSMFVVYLTATGQVNPKEHPVKOEIRIRVYNNRVKEITD 102  
DB 61 YTLNSMFVVYLTATGQVNPKEHPVKOEIRIRVYNNRVKEITD 102  
  
RESULT 9  
AB054730  
ID AB054730 standard; protein; 54 AA.  
XX  
AC AB054730;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Human genome derived single exon protein #964.  
XX  
KM Human; gene expression; single exon probe; microarray;  
XX alternative splicing event; genomic alteration.  
XX  
OS Homo sapiens.  
XX  
PN US2003194704-A1.  
XX  
PD 16-OCT-2003.  
XX  
PF 03-APR-2002; 2002US-00029386.  
XX  
PR 03-APR-2002; 2002US-00029386.  
XX  
PA (PENN/) PENN S G.  
PA (RANK/) RANK D R.  
PA (HANZ/) HANZEL D K.  
PI Penn SG, Rank DR, Hanzel DK;  
XX  
DR WPI; 2004-119264/12.  
XX  
PT New human genome-derived single exon nucleic acid probes useful for human  
PT gene expression analysis, for identifying or characterizing alternative  
PT splicing events, for assessing genomic alterations or as tools for  
PT surveying tissues.  
XX  
PS Claim 45; SEQ ID NO 28364; 80bp; English.  
XX  
CC The invention relates to a nucleic acid probe for measuring human gene  
CC expression, comprising any of the 27,400 fully defined nucleotide  
CC sequences in the specification, or their complements or fragments, and  
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences  
CC fully defined in the specification. The probe is a single exon probe that  
CC hybridizes under high stringency conditions to a nucleic acid molecule  
CC expressed in human cells or tissues. Also included are a spatially-  
CC addressed set of single exon nucleic acid probes for measuring human  
CC gene expression (comprising a plurality of single exon nucleic acid  
CC probes cited above, where each of the plurality of probes is separately





PR	21-JUL-1999	99US-0144814P
PR	21-JUL-1999	99US-0145086P
PR	21-JUL-1999	99US-0145088P
PR	22-JUL-1999	99US-0145087P
PR	22-JUL-1999	99US-0145087P
PR	22-JUL-1999	99US-0145089P
PR	22-JUL-1999	99US-0145192P
PR	22-JUL-1999	99US-0145145P
PR	23-JUL-1999	99US-0145218P
PR	23-JUL-1999	99US-0145224P
PR	23-JUL-1999	99US-0145276P
PR	26-JUL-1999	99US-0145913P
PR	27-JUL-1999	99US-0145918P
PR	27-JUL-1999	99US-0145919P
PR	28-JUL-1999	99US-0145951P
PR	02-AUG-1999	99US-0146386P
PR	02-AUG-1999	99US-0146388P
PR	02-AUG-1999	99US-0146389P
PR	03-AUG-1999	99US-0147008P
PR	04-AUG-1999	99US-0147204P
PR	05-AUG-1999	99US-0147192P
PR	05-AUG-1999	99US-0147950P
PR	09-AUG-1999	99US-0147955P
PR	10-AUG-1999	99US-0148171P
PR	11-AUG-1999	99US-0148319P
PR	12-AUG-1999	99US-0148341P
PR	13-AUG-1999	99US-0148565P
PR	13-AUG-1999	99US-0148684P
PR	16-AUG-1999	99US-0149368P
PR	17-AUG-1999	99US-0149175P
PR	18-AUG-1999	99US-0149426P
PR	20-AUG-1999	99US-0149723P
PR	20-AUG-1999	99US-0149723P
PR	22-AUG-1999	99US-0149902P
PR	23-AUG-1999	99US-0149929P
PR	23-AUG-1999	99US-0150530P
PR	25-AUG-1999	99US-0150566P
PR	26-AUG-1999	99US-0150884P
PR	27-AUG-1999	99US-0151065P
PR	27-AUG-1999	99US-0151066P
PR	27-AUG-1999	99US-0151080P
PR	30-AUG-1999	99US-0151030P
PR	31-AUG-1999	99US-0151388P
PR	01-SEP-1999	99US-0151530P
PR	07-SEP-1999	99US-0152363P
PR	10-SEP-1999	99US-0153070P
PR	13-SEP-1999	99US-0153788P
PR	15-SEP-1999	99US-0154018P
PR	16-SEP-1999	99US-0154039P
PR	20-SEP-1999	99US-0154779P
PR	22-SEP-1999	99US-0155139P
PR	22-SEP-1999	99US-0155486P
PR	24-SEP-1999	99US-0155659P
PR	28-SEP-1999	99US-0156458P
PR	29-SEP-1999	99US-0156596P
PR	04-OCT-1999	99US-0157117P
PR	05-OCT-1999	99US-0157753P
PR	06-OCT-1999	99US-0157665P
PR	07-OCT-1999	99US-0158029P
PR	08-OCT-1999	99US-0158932P
PR	12-OCT-1999	99US-0158369P
PR	13-OCT-1999	99US-0158263P
PR	13-OCT-1999	99US-0159294P
PR	13-OCT-1999	99US-0159295P
PR	14-OCT-1999	99US-0159329P
PR	14-OCT-1999	99US-0159330P
PR	14-OCT-1999	99US-0159331P
PR	14-OCT-1999	99US-0159637P
PR	14-OCT-1999	99US-0159687P

PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161923P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.

Query Match	21.6%;	Score 154.5;	DB 3;	Length 217;
Best Local Similarity	30.6%;	Pred. No. 1.2e-07;		
Matches	41;	Conservative	30;	Mismatches 46;
			Indels	17;
			Gaps	3;

```
Oy      22 ESSICAVDD-----MIKTMMAVSRENLEOKLDPEQAKVDLSVATYTLNSMFWYL 71  
         ||::||:           |:::||:|::|||   :   ::::|  
Db      18 ESALTAVNQTILAYLKELKPQEOMLTIAEPVELAAMQPLORAKMHLLAEATTTLTYELRL 77
```

```

0y      72 ATGCVNPKKHPRVKQELERIRVYNNRVEKELTDKKK----AAKLDRGAASRFLKALWEPK 126
      ||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||
Db      78 RCTGVDPDRDHRAVSEIERINVYREKFKQKVDQSKRLRPITVLNKRQATATRFIEHSL--PD 135

```

```

QY      127 RKSTPKVANKGSK 140
          ||| : : ||
Db      136 LTSTQKQSRDLK 149

```

RESULT 11  
AAG24127

AC AAG24127;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 27683.

KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.

*Arabidopsis thaliana*.

EP1033405-A2

PD 06-SEP-2000.  
yy

PF 25-FEB-2000; 2000EP-00301439.  
XX

PR	25-FEB-1999;	99US-0121825P.
PR	05-MAR-1999;	99US-0123180P.

PR	03-MAR-1999;	99US-0123548P.
PR	23-MAR-1999;	99US-0125788P.

PR 29-MAR-1999; 99US-0126785P.

PR 06-APR-1999; 99US-0128234P.

PR 16-APR-1999; 99US-0129845P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131444P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132865P.  
PR 14-MAY-1999; 99US-014256P.  
PR 14-MAY-1999; 99US-014218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 18-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 14-JUN-1999; 99US-0138847P.  
PR 16-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 24-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144335P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 23-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 06-AUG-1999; 99US-0147430P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148655P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 23-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 25-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 31-AUG-1999; 99US-0151303P.  
PR 01-SEP-1999; 99US-0151438P.  
PR 07-SEP-1999; 99US-0151930P.  
PR 10-SEP-1999; 99US-0152363P.  
PR 13-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 16-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0157865P.  
PR 08-OCT-1999; 99US-0158029P.  
PR 12-OCT-1999; 99US-0158232P.  
PR 13-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 14-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.

PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 26-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 21.6%; Score 154.5; DB 3; Length 217;  
Best Local Similarity 30.6%; Pred. No. 1.2e-07;  
Matches 41; Conservative 30; Mismatches 46; Indels 17; Gaps 3;

QY 22 ESSIGAVDD-----MLKTMMAVSRLNLLQKLDPEQAKVDLYSATLNSMFWYVL 71

DB 18 ESAIEAVNQVLAVYKEKLPQLEOM/LTLAPPEVLAMQPLQRAKTMHLLAEATTTLYEIRL 77

QY 72 ATQGVNPKHEPVKQELERIRYVMKRVKEITDKK-----AKLDRGAASRFVKALWEPK 126

DB 78 RCTGVDDDDHVRVKEIERIRINYREKFKQCVDSKGPLRPPTVLRQAKRTIEHSL--PD 135

QY 127 RKSTPKVANKGSK 140

DB 136 LTSTQKOSIRDLK 149

RESULT 12

AAG49050 ID AAG49050 standard; protein; 253 AA.

XX AAG49050;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 62012.

KW Protein identification, signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

XX Arabidopsis thaliana.

OS EPI033405-A2.

PN 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-2000; 2000EP-00301439.

PR 05-MAR-1999; 99US-0121825P.

PR 09-MAR-1999; 99US-0123180P.

PR 23-MAR-1999; 99US-0123548P.

PR 25-MAR-1999; 99US-0125788P.

PR 29-MAR-1999; 99US-0126264P.

PR 01-APR-1999; 99US-0127452P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132464P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132853P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135114P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136382P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143645P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.

```
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 02-AUG-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 13-AUG-1999; 99US-0148685P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149426P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150565P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151308P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156488P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158322P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0158933P.  
PR 13-OCT-1999; 99US-0158949P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.
```

---

```
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159684P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160880P.  
PR 22-OCT-1999; 99US-0160881P.  
PR 22-OCT-1999; 99US-0161040P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161924P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.
```

Query Match  
Best Local Similarity 30.6%; Pred. No. 1; se-07;  
Matches 41; Conservative 30; Mismatches 46; Indels 17; Gaps 3

```
QY 22 BSSGAVND-----MKTMAVSRNELLQKLDPIEOKVDIVSAYTNSMFWYL 71  
Db 54 ESALFANVQTLAYAYKEKLPOLEQLTLAEPEVLAAQIPRAKTMHLLAETTYIELRL 113  
QY 72 ATGGVNPERHPKOELERIRVNNRVKEITDPKK-----AALUDGAASRYFKALMEPK 126  
Db 114 RCTGVDDPDHRVKSEIERIIVNVREKQKCVDSKGPRTRTTVLNQAATREFTEHSL--PD 171  
QY 127 RKSTPKXANKGSK 140  
Db 172 LRTQKOSIRDLSK 185
```

RESULT 13  
AAG24126  
ID AAG24126 standard; protein; 256 AA.  
XX AC AAG24126;  
XX DE 17-OCT-2000 (first entry)  
XX DT XX  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridization assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
XX PR 05-MAR-1999; 99US-0123180P.  
XX PR 09-MAR-1999; 99US-0123548P.  
XX PR 23-MAR-1999; 99US-0125788P.  
XX PR 25-MAR-1999; 99US-0126264P.  
XX PR 25-MAR-1999; 99US-0126785P.  
XX PR 01-APR-1999; 99US-0127462P.  
XX PR 06-APR-1999; 99US-0128234P.  
XX PR 08-APR-1999; 99US-0128714P.  
XX PR 16-APR-1999; 99US-0128714P.

PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144311P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 23-JUL-1999; 99US-0145276P.  
PR 26-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 27-JUL-1999; 99US-0145951P.  
PR 28-JUL-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148655P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0151303P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152633P.  
PR 10-SEP-1999; 99US-0153707P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157153P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158932P.  
PR 12-OCT-1999; 99US-0158936P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.



PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	23-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149928P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156485P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158022P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	23-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935

PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141267P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 16-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.

PR 20-JUL-1999; 99US-0144864P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145224P.  
PR 27-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148171P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148365P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 23-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 25-AUG-1999; 99US-0149930P.  
PR 26-AUG-1999; 99US-0150566P.  
PR 27-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 31-AUG-1999; 99US-0151303P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.



```

PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

```

Query Match 20.5%; Score 146.5; DB 3; Length 177;  
 Best Local Similarity 32.4%; Pred. No. 6.5e-07;  
 Matches 36; Conservative 26; Mismatches 42; Indels 7; Gaps 2;

```

Qy 35 MMASRNELLQKDPLEQAKVDVSATYTNLSMFWYIATOGVNPKEHPVKOELERIRYTM 94
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MLTLABEPEVLAAMQPLQRAKTMHLAEATTLYELRLRCTGVDPDDHVRKXEIERINYYR 60
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 95 NRWKEITDKKK-----AAKLDGGAASRFVKKALWEPKRSSTPKVANKGSK 140
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 EKQKCVDSQSGPFRPTTVLANROAATRFIEHSL--PDLTSTQKOSIRDLSK 109
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: January 4, 2005, 13:42:28  
 Job time : 70 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 4, 2005, 11:20:50 ; Search time 16.5 Seconds  
(without alignments)  
822.216 Million cell updates/sec

Title: US-09-701-618A-4  
Perfect score: 715

Sequence: 1 MAGEENEDYPIVHESLTA.....LMEPKRSTPKVANKGSKH 141

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 263416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	142	19.9	133	2 T41115	sun-cor hormone re
2	100	14.0	184	2 S46808	hypothetical prote
3	86.5	12.1	508	2 S73430	glycerol kinase gl
4	84.5	11.8	1249	2 H71404	hypothetical prote
5	83.5	11.7	1055	2 H64577	type I restriction
6	82.5	11.5	306	2 T27985	hypothetical prote
7	82.5	11.5	2285	2 T12796	probable transgluc
8	82	11.5	530	2 G71157	hypothetical prote
9	81	11.3	207	2 S70534	bbk2.10 protein pr
10	81	11.3	207	2 S70534	bbk2.10 protein pr
11	81	11.3	478	2 CS9095	hypothetical prote
12	80.5	11.3	1133	2 T22976	hypothetical prote
13	80	11.2	317	2 T05528	hypothetical prote
14	80	11.2	992	2 S54396	protein L precursor
15	79	11.0	224	2 E64542	hypothetical prote
16	79	11.0	872	2 C72042	conserved hypotet
17	79	11.0	872	2 D86581	CT619 hypothetical
18	78.5	11.0	250	2 T22093	hypothetical prote
19	78.5	11.0	415	1 A48359	glutathyl-tRNA redu
20	78	10.9	550	2 B71441	probable myosin II
21	77	10.8	304	2 F72346	M-related protein
22	77	10.8	444	2 S64912	probable membrane
23	77	10.8	478	2 S25821	transposase - Bac
24	76.5	10.7	649	2 C75112	hypothetical prote
25	76.5	10.7	1134	2 D75014	hypothetical prote
26	76.5	10.7	1156	2 B70356	chromosome assemb
27	76.5	10.7	1708	2 AE1866	WD-40 repeat prote
28	76.5	10.7	1992	2 A47297	myosin heavy chain
29	76	10.6	471	2 S76021	probable trigger f

30	76	10.6	2415	1 A33733	spectrin alpha cha
31	75.5	10.6	117	2 T14973	hypothetical prote
32	75.5	10.6	604	2 T19682	hypothetical prote
33	75.5	10.6	629	2 F84428	probable myosin he
34	75.5	10.6	635	2 D86265	hypothetical prote
35	75.5	10.6	648	2 G72279	ATP-dependent DNA
36	75.5	10.6	2094	2 S31124	tpi protein - huma
37	75	10.5	295	2 AB2464	tRNA delta-2-isope
38	75	10.5	755	2 T20320	hypothetical prote
39	75	10.5	791	2 T20815	hypothetical prote
40	74.5	10.4	326	2 C64480	DNA repair protein
41	74.5	10.4	375	2 T32251	hypothetical prote
42	74.5	10.4	441	2 B71816	hypothetical prote
43	74.5	10.4	481	2 T28900	hypothetical prote
44	74.5	10.4	670	2 D86457	hypothetical prote
45	74.5	10.4	970	2 D59435	Gem-interacting pr

## ALIGNMENTS

RESULT 1  
T41115  
sun-cor hormone receptor repression complex homolog - fission yeast (Schizosaccharomyces  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T41115  
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z21969  
A:Accession: T41115  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-133 <L>N>  
A:Cross-references: UNIPROT:O74469; EMBL:AL031540; PDB:CAA20781.1; GSPDB:GN00068; SPDB:  
A:Experimental source: strain 972h-; cosmid cl739  
C:Genetics:  
A:Gene: SPDB:SPCC1739.07  
A:Map position: 3

Query Match 19.9%; Score 142; DB 2; Length 133;  
Best Local Similarity 28.6%; Pred. No. 3.8e-05;  
Matches 38; Conservative 24; Mismatches 57; Indels 14; Gaps 3;

QY 13 EIHESLTALSSLCAYDMLKTM-AVSRNELQKLDPLEQAVDLSAYTLNSMFVYL 71  
DB 4 EYSELPERLNKQOLDNEDVLKPKDASIPLEAGKSELEQAKLYTMSYAINSTLYSFY 63  
QY 72 ATGGVNPKEHPVQOELERIVYNNRV---KEITDKKAAKLDGGAASRVKALMEPKR 127  
DB 64 KLNGIDASERPVOEQLQRYNYSIKIQAEKNVNPTEAVNTSNAIS-----SSS 114  
QY 128 KSTPKVANKGSK 140  
DB 115 SNRPVAKDAATR 127

RESULT 2  
S46808  
hypothetical protein YHR081w - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C>Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jul-2004  
C:Accession: S46808  
R:Favell, T.  
submitted to the EMBL Data Library, June 1994  
A:Description: The sequence of S. cerevisiae cosmid 9205.  
A:Reference number: S46795  
A:Accession: S46808  
A:Molecule type: DNA  
A:Residues: 1-184 <F>V>  
A:Cross-references: UNIPROT:P38801; EMBL:U10556; NID:g500825; PDB:g500829; GSPDB:GN00008,  
C:Genetics:  
A:Gene: MIPS:YHR081w



submitted to the EMBL Data Library, November 1995  
 A/Description: The sequence of *C. elegans* cosmid ZK721.

A/Reference number: Z20450

A/Accession: F27985

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1306 <ML>

A/Cross-references: UNIPROT:Q9GVP1; EMBL:U40951; PIDN:AAA1735.1; CESP:ZK721.2

C/Genetic8:

A/Gene: CESP:ZK721.2

A/Intons: 37/3; 64/3; 114/3; 205/2; 252/3; 296/1

C/Superfamily: tropenin 1

Query Match 11.5%; Score 82.5; DB 2; Length 306;

Best Local Similarity 26.0%; Pred. No. 11;

Matches 40; Conservative 27; Mismatches 62; Indels 25; Gaps 7;

QY 2 AGEMNEDYVEIHESLTALSSIGAVDMLKTMMAVSRLNLOKLDPLEQAKV 55

DB 121 AADDLKROQLKEGERKALADRTISLPVDSID--KGOLEKIYNDMLRTGLEEKY 178

QY 56 DL--VSAYT--LNSMFWYLATQGVN-KEHPVQOELERIRVYNNRVEITDKKAAKL 109

DB 179 DINVVSQTEAEINSL-----TIEVNDLRGKFPKPSLKVKSKYDNKFKSGESKAGTKE 232

QY 110 DRGAASRFVKALME-----PKRSTPRKANCK 138

DB 233 DFRNLKIVKDDWEALVNVKKDKDKPDWSKCK 266

## RESULT 7

probable transglycoylase - *Bacillus subtilis* phage SPBc2

C/Species: *Bacillus subtilis* phage SPBc2

C/Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004

C/Accession: T12796; A69911

R/Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Manuel, C.; Karamata, D.

submitted to the EMBL Data Library, August 1997

A/Description: The complete nucleotide sequence of the *Bacillus subtilis* SPBc2 prophage

A/Reference number: Z17583

A/Accession: T12796

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 12285 <LAZ>

A/Cross-references: UNIPROT:O64046; EMBL:AF020713; NID:g3025478; PID:g3025510; PIDN:AACT

R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Aevedo, V.; Berter

C/ Bron, S.; Brouillet, S.; Bruneel, C.V.; Caldwell, B.; Capuani, V.; Carter, N.M.; Chu

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrarini, E.

Nature 390, 249-256, 1997

A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallier

leech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holappell, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli

Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,

A/Authors: Schleich, S.; Schroter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Seron

ekensch, M.; Tanakoshi, A.; Tanaka, T.; Terstegen, F.; Tognoni, A.; Toato, V.; Uchiyama,

T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A/Reference number: A69580; MUID:98044033; PMID:9384377

A/Accession: A69911

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-2285 <KUN>

A/Cross-references: GB:299115; GB:AL009126; NID:g2634478; PIDN:CAB14053.1; PID:el183582;

A/Experimental source: strain 168

C/Genetic8:

A/Gene: yomT

C/Genetic8:

A/Gene: yomT

C/Genetic8:

A/Gene: yomT

C/Genetic8:

A/Gene: yomT

C/Genetic8:

A/Gene: yomT

C/Genetic8:

A/Gene: yomT

C/Genetic8:

A/Gene: yomT

QY 9 DYVEIHESLTALSSIGAVDMLKTMMAVSRLNLOKLDPLEQAKVDLSAY----- 61

DB 1769 DQKKAIVQAKKIQQKQVNNIQKEIKTKNAKORALQELKQAKLDLSVQDQVRELQ 1828

QY 62 -----TLNSMFWYLATQGVN-KEHPVQOELERIRVYNNRVEITDKKAA 107

DB 1829 KQLVQSKVDLTAKSIESSSKTQOKIKDVONKISMEEDBDKVKYYSKQIKLQOQKKEA 1888

QY 108 KLDGASRFVKALMEPR--KSTPKV 133

DB 1889 K-----KYIKO-LLEOKKAKAGFPDI 1908

## RESULT 8

hypothetical protein PH0462 - *Pyrococcus horikoshii*

C/Species: *Pyrococcus horikoshii*

C/Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 09-Jul-2004

C/Accession: G71157

R/Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin

M.; Ohtuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi,

DNA Res. 5, 55-76, 1998

A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A/Reference number: A71000; MUID:98344137; PMID:9679194

A/Accession: G71157

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-530 <KAW>

A/Cross-references: UNIPROT:O58213; GB:AP000002; NID:g3236129; PIDN:BAA29548.1; PID:g3256

A/Experimental source: strain OT3

A/Note: this accession replaces an interim accession for a sequence replaced by Genbank

C/Genetic8:

A/Gene: PH0462

Query Match 11.5%; Score 82; DB 2; Length 530;

Best Local Similarity 25.9%; Pred. No. 23;

Matches 35; Conservative 26; Mismatches 34; Indels 40; Gaps 8;

QY 10 YPVEIHESLTALSSIGAV-----DDMLKTMMAVSRLNLOKLDPLEQAKVD 56

DB 401 YP-KIYQETVFGKSLVALKRPSPGHAHEDLTREYLSITKALFNHLSKSLDPSTFQVR 459

QY 57 LVAYATLNSM-----FWYLATQGVNKEHPVQOELERIRVYNNRVEITDKKAAKL 109

DB 460 IM-----LNSMFWYLATQGVNKEHPVQOELERIRVYNNRVEITDKKAAKL 503

QY 110 DRGAASRFVKALME 124

DB 504 ---AFYKTFYRDIWE 515

## RESULT 9

bhK2.10 protein precursor - Lyme disease spirochete

C/Species: *Borrelia burgdorferi* (Lyme disease spirochete)

C/Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 09-Jul-2004

C/Accession: S70534

R/Atkins, D.R.; Porcella, S.F.; Popova, T.G.; Shevchenko, D.; Baker, S.I.; Li, M.; Norgart

Mol. Microbiol. 18, 507-520, 1995

A/Title: Evidence for in vivo but not in vitro expression of a *Borrelia burgdorferi* outer

A/Reference number: S70531; MUID:96342380; PMID:8748034

A/Accession: S70534

A/Status: preliminary; nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-207 <AKI>

A/Cross-references: UNIPROT:Q44723; EMBL:U18292; NID:g3309519; PIDN:AAAC26095.1; PID:g8996((

C/Genetic8:

A/Gene: bhK2.10

C/Genetic8:

A/Gene: bhK2.10

C/Genetic8:

A/Gene: bhK2.10

C/Genetic8:

A/Gene: bhK2.10

C/Genetic8:

A/Gene: bhK2.10

C/Genetic8:

A/Gene: bhK2.10

C/Genetic8:

A/Gene: bhK2.10

C/Genetic8:

```
Query Match      11.3%; Score 81; DB 2; Length 207;
Best Local Similarity 25.2%; Pred. No. 9.5;
Matches 37; Conservative 25; Mismatches 53; Indels 32; Gaps 5;

QY      2 AGEENMEDYVPEIHESLTLAESLGAVDMLKTMMAVSNELTQLDLPBQAKVDLVSA 61
DB      25 SGGDIQNVKEKEGFLAKKEELVGLKKGSEAYLKVBEIMQADRPQVQAEQV---- 80
QY      62 TLNSMFWVYLATQGV--NP--KEHPVQDELRIRVYNNRVKEITDKKKAALDRGAASRF 117
DB      81 -----AAGVFEDPGLKEKLEKEIEEL-----KELKDSKKTKEDEKKELEE 122
QY      118 VKKALMEPR-----KSTPRVANKG 138
DB      123 AKQKLEPRKQVESYVENTDKVKNQK 149

RESULT 10
S70533
btk2.10 protein precursor - Lyme disease spirochete
C/Species: Borrelia burgdorferi (Lyme disease spirochete)
C/Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
R/Accession: S70533
R/Author: D.R.; Porcella, S.F.; Popova, T.G.; Shevchenko, D.; Baker, S.I.; Li, M.; Norgaard, M.; Microbiol. 18, 507-520, 1995
A/Title: Evidence for in vivo but not in vitro expression of a Borrelia burgdorferi outer surface protein F ospF
A/Reference number: S70531; PMID:96342380; PMID:8748034
A/Accession: S70533
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-207 <AKI>
A/Cross-references: UNIPROT:Q44733; EMBL:U19105; NID:9896039; PIDN:AA840878.1; PID:989603
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1994
A/Genes: btk2.10
C/Superfamily: outer surface protein F ospF
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-207/Product: btk2.10 protein #status predicted <MAT>

Query Match      11.3%; Score 81; DB 2; Length 207;
Best Local Similarity 25.2%; Pred. No. 9.5;
Matches 37; Conservative 25; Mismatches 53; Indels 32; Gaps 5;

QY      2 AGEENMEDYVPEIHESLTLAESLGAVDMLKTMMAVSNELTQLDLPBQAKVDLVSA 61
DB      25 SGGDIQNVKEKEGFLAKKEELVGLKKGSEAYLKVBEIMQADRPQVQAEQV---- 80
QY      62 TLNSMFWVYLATQGV--NP--KEHPVQDELRIRVYNNRVKEITDKKKAALDRGAASRF 117
DB      81 -----AAGVFEDPGLKEKLEKEIEEL-----KELKDSKKTKEDEKKELEE 122
QY      118 VKKALMEPR-----KSTPRVANKG 138
DB      123 AKQKLEPRKQVESYVENTDKVKNQK 149

RESULT 11
CS9095
Hypothetical protein pXOI-35 - Bacillus anthracis virulence plasmid pXOI
C/Species: Bacillus anthracis
C/Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 09-Jul-2004
R/Accession: CS9095; G59095
R/Author: K.R.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler, J.; Bacteriol. 181, 6509-6515, 1999
A/Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harboring a reference number: A59091; PMID:99445463; PMID:10515943
A/Accession: CS9095
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-478 <OKI>
A/Cross-references: UNIPROT:Q9X306; GB:AF065404; NID:94894216; PIDN:AA032339.1; PID:948942
A/Experimental source: strain Sterne
A/Note: similar to transposase for insertion sequence element IS231 (478 aa), Bacillus
```

```
A/Accession: G59095
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 77-398, 'Q', 400-401 <OK2>
A/Cross-references: GB:AF065404; NID:94894216; PIDN:AA032343.1; PID:94894255
A/Experimental source: strain Sterne
A/Note: similar to transposase for insertion sequence element IS231 (478 aa), Bacillus
A/Genes: pXOI-35; pXOI-39
C/Superfamily: transposase IS231

Query Match      11.3%; Score 81; DB 2; Length 478;
Best Local Similarity 30.2%; Pred. No. 25;
Matches 35; Conservative 13; Mismatches 46; Indels 22; Gaps 5;

QY      44 LQKLDPLEQAKVDLVSAVTLNSMFWV-----YLATQGVNPKHEHPVQDELRIRVYNN-- 95
DB      204 LDDLDQMDQGVYVYISRLKLNMMVYIKMFPEYFRNGIVKKOSQYIKVDLEHI---NMVTL 260
QY      96 ---RVKEITDK--KKAALDRGAASRFVKKALMEPRKSTPRVANKG-----KSK 140
DB      261 EPQGVYEITDAYIGKKKLTFTVYIYRLTEKOLRRKKKQVYTESKKGITYSKSK 316

RESULT 12
T22976
hypothetical protein F59A2.6 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
R/Accession: T22976; T23157
R/Author: J.
A/Title: submitted to the EMBL Data Library, June 1994
A/Reference number: Z19645
A/Accession: T22976
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-1133 <WIL>
A/Cross-references: UNIPROT:Q21022; EMBL:Z34801; PIDN:CAA64332.1; GSPDB:GN00021; CESP:F5
A/Experimental source: clone F59A2
R/Burton, J.
A/Status: submitted to the EMBL Data Library, October 1995
A/Reference number: Z19700
A/Accession: T23157
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-1133 <WIL>
A/Cross-references: EMBL:Z66514; PIDN:CAA91344.1; GSPDB:GN00021; CESP:F59A2.6
A/Experimental source: clone K01A11
A/Genes: CESP:F59A2.6
A/Map position: 3
A/Introns: 13/2; 43/3; 107/3; 413/3; 492/3; 567/3; 635/3; 710/3; 738/3; 795/2; 1008/3; 1055/28

Query Match      11.3%; Score 80.5; DB 2; Length 1133;
Best Local Similarity 26.1%; Pred. No. 73;
Matches 42; Conservative 23; Mismatches 53; Indels 43; Gaps 7;

QY      1 MAGEENM--EDYVPEIHESLTLAESLGAVDMLKTMMAV-----SNNELT-- 44
DB      361 LAGELNMAEDLKVYBEKHGTIQAGQALDPAKEEVVYLKEQLERAQSALESSEGLASS 420
QY      45 QKLDLPBQAKVDLVSAV-----TLNSMFWVYLAT--QGVNPKHEHPVQDELRIRVYNN 95
DB      421 QKADKQDELEKELQNAQKRSSSELETINEMVRSILATLNSNSTETITLKQLETLT----- 475
QY      96 RVKEITDKKKAALDRGAASRFVKKALMEPRKSTPRVANK 136
DB      476 -----DKELQRR-----QOTEKALTEINVLVLSIAEK 503

RESULT 13
T05528
```

hypothetical protein F13M23.210 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004

C/Accession: T05528

R/Levan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohnes, J.; Mewes, H.W.; Mayer, K.F.

submitted to the Protein Sequence Database, February 1999

A/Reference number: 215419

A/Accession: T05528

A/Molecule type: DNA

A/Residues: 1-317 <BEV>

A/Cross-references: UNIPROT:Q9SM19; EMBL:AL035523

A/Experimental source: cultivar Columbia; BAC clone F13M23

C/Genetics:

A/Map position: 4

A/Intons: 23/3; 48/3; 98/3; 119/1; 157/1; 196/3; 220/1; 233/3; 276/1

A/Note: F13M23.210

Query Match 11.2%; Score 80; DB 2; Length 317;

Best Local Similarity 24.2%; Pred. No. 19;

Matches 31; Conservative 27; Mismatches 46; Indels 24; Gaps 6;

QY 16 ESITALESSLGAVDDML--KTMMASR-----NELQKLDPLEQAKVDLSAYTLNS 65

Db 180 ESMIAVETGLREITASLKVEDAVMLALSRVQTNVAVQAVTELSDHEDILFKKAWTL-- 237

QY 66 MFWNYLATQGVNPKHPVKEHVKOEL--ERIRVYNNRVKEITDKKKAAKLDRG--AASRFVKK 120

Db 238 YFM-----KRAKHGESDIDAEERLQFWINRLGQSSSHDAIDVERGKEKLKGLIEQ 290

QY 121 ALWEPRKK 128

Db 291 QLMETSRK 298

RESULT 14

S54396

protein L precursor - Peptostreptococcus magnus (strain 3316)

C/Species: Peptostreptococcus magnus

A/Variety: Strain 3316

C/Date: 27-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 09-Jul-2004

C/Accession: S54396

R/Murphy, J.P.; Duggieby, C.J.; Atkinson, M.A.; Trowern, A.R.; Atkinson, T.; Goward, C.F.

Mol. Microbiol. 12, 911-920, 1994

A/Title: The functional units of a peptostreptococcal protein L.

A/Reference number: S54396; MUID:95020613; PMID:7934898

A/Accession: S54396

A/Status: preliminary; nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-992 <MUR>

A/Cross-references: UNIPROT:Q51918; EMBL:L04466; NID:g150673; PIDN:AAA67503.1; PID:g1506

Query Match 11.2%; Score 80; DB 2; Length 992;

Best Local Similarity 23.6%; Pred. No. 69;

Matches 29; Conservative 32; Mismatches 42; Indels 20; Gaps 5;

QY 9 DYVEIHESITALESSLGAVDDMLKTMMASRNELLQKLDPLEQ-----AKVDLSA 60

Db 47 DYHGDVSDVDPVEE---IDEALAKALAEKERTAKGHIDSLNLSSTAKKLANDIDSA 103

QY 61 YTLNSMFWNYLATQGVNPKHPVKEHVKOEL-----RVYNNRVKEITDK--KKAAXLDRG 112

Db 104 YTLNAINDI--VARADVNERKTAEKEAEKLAATAKETAKKHIDELKHLADTKKELAKRDID 162

QY 113 AAS 115

Db 163 SAT 165

RESULT 15

E64542

hypothetical protein HP0181 - Helicobacter pylori (strain 26695)

C/Species: Helicobacter pylori

C/Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 09-Jul-2004

C/Accession: E64542

R/Tomb, J.F.; White, O.; Kervage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Giodok, A.; McKenney,

son, J.D.; Kelley, J.M.; Cooton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watney, L.

Nature 388, 539-547, 1997

A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A/Reference number: A64520; MUID:97394467; PMID:9252185

A/Accession: E64542

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-224 <TOM>

A/Cross-references: UNIPROT:Q24983; GB:AE000538; GB:AE000511; NID:g2313263; PIDN:AD0725;

C/Genetics:

A/Start codon: GTG

Query Match 11.0%; Score 79; DB 2; Length 224;

Best Local Similarity 24.8%; Pred. No. 15;

Matches 34; Conservative 27; Mismatches 54; Indels 22; Gaps 6;

QY 12 VEIHESITALESSLGAVDDMLKTMM-----AVSRNELLQKLDPLEQAKVDLSAY--T 62

Db 82 VLVFSGGIIDKALGFIFSCLTFTFLVSLFLVALSKVEVKNANAYLQEK-----SAFFST 137

QY 63 LNSMFWNYLATQGVNPKHPVKEHVKOELERIRVYNNRVKEITDKKKAA--KLDGGAAS---- 115

Db 138 MKSVASKIMLDGKAVNEONLKNLBSMSDEVCK--KESFVKAKESFVKAMDGVESLKEK 196

QY 116 -REVKKALWEPRKRSKP 131

Db 197 AKDLPRKMLDPRKANQTP 213

Search completed: January 4, 2005, 13:45:15  
Job time: 18.5 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 5, 2005, 13:12:38 ; Search time 68 Seconds

(without alignments)  
1473.842 Million cell updates/sec

Title: US-09-701-618a-4

Perfect score: 715  
Sequence: 1 MAGEENNEDVPVEIHSLTA.....LMPEKRSKTPKVNKSKSKH 141

## Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US09701618/runat\_05012005\_085839\_24463/app\_query.fasta\_1.654  
-DB=Issued Patents NA -OPMT=fastap -SUPFIX=rml -MINMATCH=0.1 -LOOPCTL=0  
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdl  
-LIST=45 -DOCALLIN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09701618.OCGN.1.1.141@runat\_05012005\_085839\_24463 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	485	67.8	380	US-09-513-999C-1235	Sequence 1235, Ap
2	113	15.8	651	US-09-248-796A-841	Sequence 841, App
3	85.5	12.0	885	US-09-134-000C-332	Sequence 332, App
4	84	11.7	819	US-09-248-796A-4954	Sequence 4954, Ap
5	82.5	11.5	7100	US-09-308-375-1	Sequence 1, Appl
6	82	11.5	924	US-09-107-532A-3386	Sequence 3386, Ap
7	80.5	11.3	1407	US-09-248-796A-13211	Sequence 13211, A
8	80.5	11.3	1428	US-09-248-796A-1115	Sequence 1115, Ap
9	80.5	11.3	1710	US-08-747-221B-33	Sequence 33, Appl
10	80.5	11.3	1710	US-09-005-051-33	Sequence 33, Appl
11	80.5	11.3	1710	US-09-403-942F-33	Sequence 33, Appl
12	80.5	11.3	1785	US-08-747-221B-34	Sequence 34, Appl

C 13	80.5	11.3	1785	3	US-08-747-221B-35	Sequence 35, Appl
C 14	80.5	11.3	1785	3	US-09-005-051-34	Sequence 34, Appl
C 15	80.5	11.3	1785	3	US-09-005-051-35	Sequence 35, Appl
C 16	80.5	11.3	1785	4	US-09-403-942F-34	Sequence 34, Appl
C 17	80.5	11.3	1785	4	US-09-403-942F-35	Sequence 35, Appl
C 18	80.5	11.3	1785	3	US-08-747-221B-33	Sequence 30, Appl
C 19	80.5	11.3	2801	3	US-08-747-221B-32	Sequence 32, Appl
C 20	80.5	11.3	2801	3	US-09-005-051-30	Sequence 30, Appl
C 21	80.5	11.3	2801	4	US-09-005-051-32	Sequence 32, Appl
C 22	80.5	11.3	2801	4	US-09-403-942F-30	Sequence 30, Appl
C 23	80.5	11.3	2801	4	US-09-403-942F-32	Sequence 32, Appl
C 24	80.5	11.3	1664976	4	US-08-916-421B-1	Sequence 1, Appl
C 25	80.5	11.3	1664976	4	US-09-692-570-1	Sequence 1, Appl
C 26	80	11.2	3279	3	US-08-446-137B-1	Sequence 1, Appl
C 27	79	11.0	123025	4	US-09-198-452A-1	Sequence 1, Appl
C 28	78.5	11.0	1710	3	US-08-747-221B-27	Sequence 27, Appl
C 29	78.5	11.0	1710	3	US-09-005-051-27	Sequence 27, Appl
C 30	78.5	11.0	1710	4	US-09-403-942F-27	Sequence 27, Appl
C 31	78.5	11.0	1788	3	US-08-747-221B-28	Sequence 28, Appl
C 32	78.5	11.0	1788	3	US-08-747-221B-29	Sequence 29, Appl
C 33	78.5	11.0	1788	3	US-09-005-051-28	Sequence 28, Appl
C 34	78.5	11.0	1788	3	US-09-005-051-29	Sequence 29, Appl
C 35	78.5	11.0	1788	4	US-09-403-942F-28	Sequence 28, Appl
C 36	78.5	11.0	1788	4	US-09-403-942F-29	Sequence 29, Appl
C 37	78.5	11.0	2836	3	US-08-747-221B-24	Sequence 24, Appl
C 38	78.5	11.0	2836	3	US-08-747-221B-26	Sequence 26, Appl
C 39	78.5	11.0	2836	3	US-09-005-051-26	Sequence 26, Appl
C 40	78.5	11.0	2836	3	US-09-005-051-24	Sequence 24, Appl
C 41	78.5	11.0	2836	4	US-09-403-942F-24	Sequence 24, Appl
C 42	78.5	11.0	2836	4	US-09-403-942F-26	Sequence 26, Appl
C 43	78	10.9	456	4	US-09-134-000C-3074	Sequence 3074, Ap
C 44	77.5	10.8	579	4	US-09-134-000C-3226	Sequence 3226, Ap
C 45	77	10.8	2433	4	US-09-248-796A-6178	Sequence 6178, Ap

## ALIGNMENTS

RESULT 1  
US-09-513-999C-1235  
Sequence 1235, Application US/09513999C  
Patent No. 6783961  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Duclet, A. Y.  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
PATENT NO. 6783961  
FILE REFERENCE: 59,US2,REG  
CURRENT FILING DATE: 2000-02-24  
CURRENT APPLICATION NUMBER: US/09/513,999C  
PRIOR FILING DATE: 1999-02-26  
NUMBER OF SEQ ID NOS: 36681  
SOFTWARE: Patent.pm  
SEQ ID NO 1235  
LENGTH: 380  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 75..380  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 9  
OTHER INFORMATION: s=g or c  
US-09-513-999C-1235  
Alignment Scores:  
Pred. No.: 9.92e-62  
Score: 485.00  
Percent Similarity: 98.04%  
Best Local Similarity: 92.16%  
Query Match: 67.83%  
Length: 380  
Matches: 94  
Conservative: 6  
Mismatch: 2  
Indels: 0

DB:	4	Gaps:	0
US-09-701-618A-4 (1-141) x US-09-513-999C-1235 (1-380)			
QY	1 MetAlaGlyGluGluMetAenGluAepTyrProValGluIleHisGluSerLeuThrAla 20		
Db	75 ATGGCAGGTGAGAAATTAATGAAGCTATCCAGTGAATTAACAGATTTTGTGACGG 134		
QY	21 LeuGluSerSerLeuGluValAlaAspAspMetLeuIleThrMetMetAlaValSerArg 40		
Db	135 TTGGAGATTCACATGGTGTGTGATGACATCTGAAGCAGCATGATGCTGTGTTCTAGA 194		
QY	41 AsnGluLeuLeuGluIleuLysLeuAspProLeuGluGlnAlaIleValAspLeuValSerAla 60		
Db	195 AATGAGTGTGTGCAGAAAGTTGGATTCACCTGAAACAGCAAAAGTGGATTTGGTTCTGCA 254		
QY	61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80		
Db	255 TACACATTAATTCATATGTTTGGGTTTATTTGGCAACCAAGAGTTATCTTAAGGAA 314		
QY	81 HisProValLysGlnIleuLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100		
Db	315 CATCCAGTAAACAGGAATTGGAAAGATCAAGATGATATATGAACAGAGTCAAGGAAATA 374		
QY	101 ThrAsp 102		
Db	375 ACAGAC 380		
RESULT 2			
US-09-248-796A-841			
/ Sequence 841, Application US/09248796A			
/ Patent No. 6747137			
/ GENERAL INFORMATION:			
/ APPLICANT: Keith weinstock et al			
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS			
/ FILE REFERENCE: 107196.132			
/ CURRENT APPLICATION NUMBER: US/09/248, 796A			
/ PRIOR FILING DATE: 1998-02-12			
/ PRIOR APPLICATION NUMBER: US 60/074, 725			
/ PRIOR FILING DATE: 1998-02-13			
/ PRIOR APPLICATION NUMBER: US 60/096, 409			
/ PRIOR FILING DATE: 1998-08-13			
/ NUMBER OF SEQ ID NOS: 28208			
/ SEQ ID NO 841			
/ LENGTH: 651			
/ TYPE: DNA			
/ ORGANISM: Candida albicans			
US-09-248-796A-841			
Alignment Scores:			
Pred. No.:	7,24e-07	Length:	651
Score:	113.00	Matches:	33
Percent Similarity:	44.17%	Conservative:	20
Best Local Similarity:	27.50%	Mismatches:	47
Query Match:	15.80%	Indels:	20
DB:	4	Gaps:	2
US-09-701-618A-4 (1-141) x US-09-248-796A-841 (1-651)			
QY	10 TyrProValGluIleHisGluSerLeuThrAlaLeuGluSerSerLeuGly----- 26		
Db	52 TATATAACATCATTAATCACAATCATCATTCAGCTTATGAATCTCATTAAGTCCATTACAA 111		
QY	27 -----AlaValAspAspMetLeuLysThrMetMetAlaValSerTrpArgSerGluLeu 44		
Db	112 AACAAACAATTAATCCGATCATGATTTCTAAATATCAACACAAATCATCAACACCTGTACC 171		
QY	45 GlnLysLeuAspProLeuGluGlnAlaIleValAspLeuValSerAlaTyrThrLeuAsn 64		
Db	172 ACTTTCATACATCCAGAGAACAAATTCAAATGATGTAATTAATTTGTCATATATTTTAAATA 231		
QY	65 SerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGluHisProValLys 84		

QY	Db	232	TCGACATTATTCAGTTATTTTAAATCCTTGAATTAATCTGATTCCTCATCCATATAA	291
QY	85	glnqluenglualrgllleargvaltyrmetasnarvallysgluile	-----	100
Db	292	ATGGAATTAATCAAGAAATCAATCATCATGATTCGATGAAAATATTTAAATATGAATA	351	
QY	101	-----Thrasplyslyslvalaialysleu	109	
Db	352	AATGCTGATACCATTAACAAGAAAGAGAAAGAAAGAAAATTTAAAGAACTG	411	
RESULT 3				
US-09-134-000C-332				
; Sequence 332, Application US/09134000C				
; Patent NO. 6617156				
; GENERAL INFORMATION:				
; APPLICANT: Lynn Doucette-Stamm et al				
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO				
; FILE REFERENCE: 032796-032				
; CURRENT APPLICATION NUMBER: US/09/134, 000C				
; PRIOR APPLICATION NUMBER: 1998-08-13				
; PRIOR FILING DATE: 1997-08-15				
; NUMBER OF SEQ ID NOS: 6812				
; SOFTWARE: PatentIn version 3.1				
; SEQ ID NO 332				
; LENGTH: 885				
; TYPE: DNA				
; ORGANISM: Enterococcus faecalis				
US-09-134-000C-332				
Alignment Scores:				
Pred. No.:	0.0126	Length:	885	
Score:	85.50	Matches:	38	
Percent Similarity:	42.76%	Conservative:	24	
Best Local Similarity:	26.21%	Mismatches:	50	
Query Match:	11.96%	Indels:	33	
DB:	4	Gaps:	6	
US-09-701-618A-4 (1-141) x US-09-134-000C-332 (1-885)				
QY	1	MetlaaglygluGlumetAsnGluaspTyrProvalGluileHlsGluSerleuThrala	20	
Db	199	ATGGCTGGTGGAAGTTTCTCGAAGATCAACAAGAGATTGAAGACGTTTAAAC---	255	
QY	21	LeugluserSerleuGlyAlavalaspaSmetleuysthrmetMelAlvalSerarg	40	
Db	256	-----ATGAATATTCCTTCATTCAAGAAATTTTAGTACACAC	294	
QY	41	AsnGluLeuLeuGlnLysLeuaspProleuGlnAlalyValaspleuValSerAla	60	
Db	295	CAAGAGCTGTATCAACAATACCCGTGTTTAGACATGACA-----GACTTAGAAAAGAGC	348	
QY	61	TyrThrleuAsnSerMetPheTrpValTyrleuAlathrglnGlyValAsnProlysglu	80	
Db	349	GCGAATTAATCGCTGATCTTAAGCAATATATTTCTGAAGCCATCAAA---AAAGAT	405	
QY	81	HisProval-----LysGlnGluLeuGlu	88	
Db	406	TATTTAATCAATACGTACAAACATCTTCACAGTTAGTTAGTCGCAAAAGAGAGCTGGAG	465	
QY	89	ArglleargValTyrMetAsnargVallysgluileThrasplyslyslvalaialys	108	
Db	466	CCATTGAAATATGATCCAA-----AAGATATTCACAAACTTATTAACGTGCACAA	516	
QY	109	LeuasparglyAlaialasergPhevalLysLys-----Ala	121	
Db	517	TTAACTCCCTTCAAGTGCCTTATCAACTTAAATAATTAACCAATTACACCGGTTTATGATGCC	576	
QY	122	LeuTrpGluProlys	126	
Db	577	CTTTGGGAAAAACAA	591	



```

1 / ZIP: 02354
2 / COMPUTER READABLE FORM:
3 / MEDIUM TYPE: CD-ROM ISO9660
4 / COMPUTER: PC
5 / OPERATING SYSTEM: <Unknown>
6 / SOFTWARE: ASCII
7 / CURRENT APPLICATION DATA:
8 / APPLICATION NUMBER: US/09/107,532A
9 / FILING DATE: 30-Jun-1998
10 / PRIOR APPLICATION DATA:
11 / APPLICATION NUMBER: 60/085,598
12 / FILING DATE: 14 May 1998
13 / APPLICATION NUMBER: 60/051571
14 / FILING DATE: July 2, 1997
15 / ATTORNEY/AGENT INFORMATION:
16 / NAME: Ariniello, Pamela Deneke
17 / REGISTRATION NUMBER: 40,489
18 / REFERENCE/DOCKET NUMBER: 40,489
19 / TELECOMMUNICATION INFORMATION:
20 / TELEPHONE: (781)893-5007
21 / TELEFAX: (781)893-8277
22 / INFORMATION FOR SEQ ID NO: 3386:
23 / SEQUENCE CHARACTERISTICS:
24 / LENGTH: 924 base pairs
25 / TYPE: nucleic acid
26 / STRANDEDNESS: double
27 / TOPOLOGY: circular
28 / MOLECULE TYPE: DNA (genomic)
29 / HYPOTHEetical: NO
30 / ANTI-SENSE: NO
31 / ORIGINAL SOURCE:
32 / ORGANISM: Enterococcus faecium
33 / FEATURE:
34 / NAME/KEY: misc_feature
35 / LOCATION: (b) LOCATION 1...924
36 / SEQUENCE DESCRIPTION: SEQ ID NO: 3386:
37 /
38 /
39 /
40 /
41 /
42 /
43 /
44 /
45 /
46 /
47 /
48 /
49 /
50 /
51 /
52 /
53 /
54 /
55 /
56 /
57 /
58 /
59 /
60 /
61 /
62 /
63 /
64 /
65 /
66 /
67 /
68 /
69 /
70 /
71 /
72 /
73 /
74 /
75 /
76 /
77 /
78 /
79 /
80 /
81 /
82 /
83 /
84 /
85 /
86 /
87 /
88 /
89 /
90 /
91 /
92 /
93 /
94 /
95 /
96 /
97 /
98 /
99 /
100 /
101 /
102 /
103 /
104 /
105 /
106 /
107 /
108 /
109 /
110 /
111 /
112 /
113 /
114 /
115 /
116 /
117 /
118 /
119 /
120 /
121 /
122 /
123 /
124 /
125 /
126 /
127 /
128 /
129 /
130 /
131 /
132 /
133 /
134 /
135 /
136 /
137 /
138 /
139 /
140 /
141 /
142 /
143 /
144 /
145 /
146 /
147 /
148 /
149 /
150 /
151 /
152 /
153 /
154 /
155 /
156 /
157 /
158 /
159 /
160 /
161 /
162 /
163 /
164 /
165 /
166 /
167 /
168 /
169 /
170 /
171 /
172 /
173 /
174 /
175 /
176 /
177 /
178 /
179 /
180 /
181 /
182 /
183 /
184 /
185 /
186 /
187 /
188 /
189 /
190 /
191 /
192 /
193 /
194 /
195 /
196 /
197 /
198 /
199 /
200 /
201 /
202 /
203 /
204 /
205 /
206 /
207 /
208 /
209 /
210 /
211 /
212 /
213 /
214 /
215 /
216 /
217 /
218 /
219 /
220 /
221 /
222 /
223 /
224 /
225 /
226 /
227 /
228 /
229 /
230 /
231 /
232 /
233 /
234 /
235 /
236 /
237 /
238 /
239 /
240 /
241 /
242 /
243 /
244 /
245 /
246 /
247 /
248 /
249 /
250 /
251 /
252 /
253 /
254 /
255 /
256 /
257 /
258 /
259 /
260 /
261 /
262 /
263 /
264 /
265 /
266 /
267 /
268 /
269 /
270 /
271 /
272 /
273 /
274 /
275 /
276 /
277 /
278 /
279 /
280 /
281 /
282 /
283 /
284 /
285 /
286 /
287 /
288 /
289 /
290 /
291 /
292 /
293 /
294 /
295 /
296 /
297 /
298 /
299 /
300 /
301 /
302 /
303 /
304 /
305 /
306 /
307 /
308 /
309 /
310 /
311 /
312 /
313 /
314 /
315 /
316 /
317 /
318 /
319 /
320 /
321 /
322 /
323 /
324 /
325 /
326 /
327 /
328 /
329 /
330 /
331 /
332 /
333 /
334 /
335 /
336 /
337 /
338 /
339 /
340 /
341 /
342 /
343 /
344 /
345 /
346 /
347 /
348 /
349 /
350 /
351 /
352 /
353 /
354 /
355 /
356 /
357 /
358 /
359 /
360 /
361 /
362 /
363 /
364 /
365 /
366 /
367 /
368 /
369 /
370 /
371 /
372 /
373 /
374 /
375 /
376 /
377 /
378 /
379 /
380 /
381 /
382 /
383 /
384 /
385 /
386 /
387 /
388 /
389 /
390 /
391 /
392 /
393 /
394 /
395 /
396 /
397 /
398 /
399 /
400 /
401 /
402 /
403 /
404 /
405 /
406 /
407 /
408 /
409 /
410 /
411 /
412 /
413 /
414 /
415 /
416 /
417 /
418 /
419 /
420 /
421 /
422 /
423 /
424 /
425 /
426 /
427 /
428 /
429 /
430 /
431 /
432 /
433 /
434 /
435 /
436 /
437 /
438 /
439 /
440 /
441 /
442 /
443 /
444 /
445 /
446 /
447 /
448 /
449 /
450 /
451 /
452 /
453 /
454 /
455 /
456 /
457 /
458 /
459 /
460 /
461 /
462 /
463 /
464 /
465 /
466 /
467 /
468 /
469 /
470 /
471 /
472 /
473 /
474 /
475 /
476 /
477 /
478 /
479 /
480 /
481 /
482 /
483 /
484 /
485 /
486 /
487 /
488 /
489 /
490 /
491 /
492 /
493 /
494 /
495 /
496 /
497 /
498 /
499 /
500 /
501 /
502 /
503 /
504 /
505 /
506 /
507 /
508 /
509 /
510 /
511 /
512 /
513 /
514 /
515 /
516 /
517 /
518 /
519 /
520 /
521 /
522 /
523 /
524 /
525 /
526 /
527 /
528 /
529 /
530 /
531 /
532 /
533 /
534 /
535 /
536 /
537 /
538 /
539 /
540 /
541 /
542 /
543 /
544 /
545 /
546 /
547 /
548 /
549 /
550 /
551 /
552 /
553 /
554 /
555 /
556 /
557 /
558 /
559 /
560 /
561 /
562 /
563 /
564 /
565 /
566 /
567 /
568 /
569 /
570 /
571 /
572 /
573 /
574 /
575 /
576 /
577 /
578 /
579 /
580 /
581 /
582 /
583 /
584 /
585 /
586 /
587 /
588 /
589 /
590 /
591 /
592 /
593 /
594 /
595 /
596 /
597 /
598 /
599 /
600 /
601 /
602 /
603 /
604 /
605 /
606 /
607 /
608 /
609 /
610 /
611 /
612 /
613 /
614 /
615 /
616 /
617 /
618 /
619 /
620 /
621 /
622 /
623 /
624 /
625 /
626 /
627 /
628 /
629 /
630 /
631 /
632 /
633 /
634 /
635 /
636 /
637 /
638 /
639 /
640 /
641 /
642 /
643 /
644 /
645 /
646 /
647 /
648 /
649 /
650 /
651 /
652 /
653 /
654 /
655 /
656 /
657 /
658 /
659 /
660 /
661 /
662 /
663 /
664 /
665 /
666 /
667 /
668 /
669 /
670 /
671 /
672 /
673 /
674 /
675 /
676 /
677 /
678 /
679 /
680 /
681 /
682 /
683 /
684 /
685 /
686 /
687 /
688 /
689 /
690 /
691 /
692 /
693 /
694 /
695 /
696 /
697 /
698 /
699 /
700 /
701 /
702 /
703 /
704 /
705 /
706 /
707 /
708 /
709 /
710 /
711 /
712 /
713 /
714 /
715 /
716 /
717 /
718 /
719 /
720 /
721 /
722 /
723 /
724 /
725 /
726 /
727 /
728 /
729 /
730 /
731 /
732 /
733 /
734 /
735 /
736 /
737 /
738 /
739 /
740 /
741 /
742 /
743 /
744 /
745 /
746 /
747 /
748 /
749 /
750 /
751 /
752 /
753 /
754 /
755 /
756 /
757 /
758 /
759 /
760 /
761 /
762 /
763 /
764 /
765 /
766 /
767 /
768 /
769 /
770 /
771 /
772 /
773 /
774 /
775 /
776 /

```

```

Db          592 CTACTTCTTCATATTCTCAGTAAGTCGACGACACTAAGCGATAAA 639
                               ::::: |||::: ::: |||||
RESULT 7
US-09-248-796A-13211      ; Sequence 13211, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Kelth Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 13211
; LENGTH: 1407
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-13211

Alignment Scores:
Pred. No.:           0.143       Length:         1407
Score:              80.50        Matches:         29
Percent Similarity: 45.79%       Conservative:    20
Best Local Similarity: 27.10%     Mismatches:     35
Query Match:        11.26%        Indels:         23
DB:                  4           Gaps:             5

US-09-701-618A-4 (1-141) x US-09-248-796A-13211 (1-1407)
Qy      3  GLYGLIGLMeTAsnGUAsPtyrProValGIuIleHISgluSerLeuThrAlaleuGIU 22
Db      793 GGGAACAAGTTGTACGAAGAATGGATGGATTGAAGTTCAGAA--TTATTGAAAATGCAC 849
               ||||| |
Qy      23 SerSerrUGlAlaValAlasp---AspMetLeuLythrMetMecrAlaValSerArgasn 41
Db      850 TTAATCTACCACCAAGCATGAGATGAGAAATTAATGATGATATGATGATGCTTACTGCCAAC 909
               ||||| |
Qy      42 -----GluLeuLeuGIuLySerLeuAspProLeuGIU 51
Db      910 TTGCAGATGTCCGTAAGTGCGAGACATTTTTTTGTCTAATAGATTGGAAACCGAAGAT 969
               ||||| |
Qy      52 GluAlaLySvAlasp-----LeuValSerAlaTyThrLeuAsnSer 65
Db      970 GCAGGTGAGTTGATTGAACAACGCAACGATATGATCAATATCATATATATGAC 1029
               ||||| |
Qy      66 Met-----PheTrpValLytleuAlaThrGInLyValAsnProLySgu 80
Db      1030 TTGATCTATGTTCAAGAAATTTTGGGATTAACCTCTCAATGTTGGATTGGTTCGGATTAT 1089
               ||||| |
Qy      81 HisProValLySGInGUleu 87
Db      1090 CAAATGTTCAAACAATACTTG 1110
               ||||| |

RESULT 8
US-09-248-796A-1115
; Sequence 1115, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Kelth Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409

```



```

/ ADDRESSEE: Heska Corporation
/ STREET: 1825 Sharp Point Drive
/ CITY: Fort Collins
/ STATE: Colorado
/ COUNTRY: USA
/ ZIP: 80525
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: WordPerfect for Windows, Version 7.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/005,051
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/747,221
/ FILING DATE: No. 6291222ember 12, 1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Vetter, Carol Talkington
/ REGISTRATION NUMBER: 37,459
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 970/493-7272
/ TELEFAX: 970/484-9505
/ INFORMATION FOR SEQ ID NO: 33:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1710 nucleotides
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..1710
/ US-09-005-051-33

Alignment Scores:
Pred. No.: 0.196 Length: 1710
Score: 80.50 Matches: 38
Percent Similarity: 38.41% Conservative: 25
Best Local Similarity: 23.17% Mismatches: 62
Query Match: 11.26% Indels: 39
DB: Gaps: 8

US-09-701-618a-4 (1-141) x US-09-005-051-33 (1-1710)
QY 4 GluGluMetAenGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23
Db 1138 AAAGAACAAAACGATGACATC-----GAGGCTATGTCACCACTAAAGGC 1182
QY 24 SerLeuGlyAlaValAspAspMetLeuLysThrMetMetAlaValSerArgAenGluLeu 43
Db 1183 GATCGTGGTATCTCCAAAGAAATTACCGTACCTTGAAAGCCATATTTTTCATGAATC 1242
QY 44 LeuGlnLys-----LeuAspProLeuGluGlnAlaValAspLeuValSer 59
Db 1243 AAAAGAAATTCACACTGTATTTGTATAGGTATACAGATGATACGTATAGTATATAA 1302
QY 60 AlaTyrThrLeu----- 63
Db 1303 AGTTATATCTTGGCCCTATGATGGGTTCTTCCAGAGATTAGTCAGTATGATTTTA 1362
QY 64 -----AsnSerMetPheTyrValTyrLeuAlaThrGlnIleValAsn--- 77
Db 1363 GGAATATCTTTTGGCAAACTCTTGGATGTTCATTTTGGAAACACGACATTTCTATA 1422
QY 78 ProLysGluHisProValLysGlnIleGluLysArg---IleArgValTyrMetAsnArg 96
Db 1423 CCGCAAGAT-----GCTATGCAAGACTCTGGAAAGAGATGTCAGATCTGACCAATTTT 1476
QY 97 ValLys-----GluIleThrAspLysLysAlaAlaLysLeuAspArgGlyAlaAla 114

```

```

Db 1477 GTAAAGATGAAAACTTACATCAACACCTGAGATGATCATGATGAT-----ACA 1527
QY 115 SerArgPheValLysLysAlaLeuTyrGluProLysArgLysSerThrProLysValAla 134
Db 1528 AAAAGACATTTTAAACGACATTTTGGGAACCATACACAGACGAAACCAATAATTTTG 1587
QY 135 AsnLysGlyLys 138
Db 1588 GACATGGGAAAA 1599

RESULT 11
US-09-403-942F-33
/ Sequence 33, Application US/09403942F
/ Patent No. 6664090
/ GENERAL INFORMATION:
/ APPLICANT: Silver, Gary M.
/ APPLICANT: Brandt, Kevin S.
/ TITLE OF INVENTION: NOVEL CARBOXYL ESTERASE NUCLEIC ACID MOLECULES, PROTEINS AND USES
/ FILE REFERENCE: FC-1-C1-PUS
/ CURRENT APPLICATION NUMBER: US/09/403,942F
/ PRIOR APPLICATION NUMBER: 2000-05-02
/ PRIOR FILING DATE: 1997-11-10
/ PRIOR APPLICATION NUMBER: PCT/US97/20598
/ PRIOR FILING DATE: 1996-11-12
/ NUMBER OF SEQ ID NOS: 76
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 33
/ LENGTH: 1710
/ TYPE: DNA
/ ORGANISM: Ctenocephalides felis
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: (1)..(1710)
/ OTHER INFORMATION:
/ US-09-403-942F-33

Alignment Scores:
Pred. No.: 0.196 Length: 1710
Score: 80.50 Matches: 38
Percent Similarity: 38.41% Conservative: 25
Best Local Similarity: 23.17% Mismatches: 62
Query Match: 11.26% Indels: 39
DB: Gaps: 8

US-09-701-618a-4 (1-141) x US-09-403-942F-33 (1-1710)
QY 4 GluGluMetAenGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23
Db 1138 AAAGAACAAAACGATGACATC-----GAGGCTATGTCACCACTAAAGGC 1182
QY 24 SerLeuGlyAlaValAspAspMetLeuLysThrMetMetAlaValSerArgAenGluLeu 43
Db 1183 GATCGTGGTATCTCCAAAGAAATTACCGTACCTTGAAAGCCATATTTTTCATGAATC 1242
QY 44 LeuGlnLys-----LeuAspProLeuGluGlnAlaValAspLeuValSer 59
Db 1243 AAAAGAAATTCACACTGTATTTGTATAGGTATACAGATGATACGTATAGTATATAA 1302
QY 60 AlaTyrThrLeu----- 63
Db 1303 AGTTATATCTTGGCCCTATGATGGGTTCTTCCAGAGATTAGTCAGTATGATTTTA 1362
QY 64 -----AsnSerMetPheTyrValTyrLeuAlaThrGlnIleValAsn--- 77
Db 1363 GGAATATCTTTTGGCAAACTCTTGGATGTTCATTTTGGAAACACGACATTTCTATA 1422
QY 78 ProLysGluHisProValLysGlnIleGluLysArg---IleArgValTyrMetAsnArg 96
Db 1423 CCGCAAGAT-----GCTATGCAAGACTCTGGAAAGAGATGTCAGATCTGACCAATTTT 1476
QY 97 ValLys-----GluIleThrAspLysLysAlaAlaLysLeuAspArgGlyAlaAla 114

```

Db 1477 GTAAAGATGGAACCTACATCAACACCTGACATGATGAT-----ACA 1527  
Qy 115 SerArgPheValIySlybAlaLeuTrpGluProLysArgIysSerThrProLysValAla 134  
Db 1528 AAAAGACATTTTAAACGACATTTTGGGACCATACACGACGAAGAACCAAAATATTG 1587  
Qy 135 AsnLysGlyLys 138  
Db 1588 GACATGGGAAAA 1599  
RESULT 12  
US-08-747-221B-34  
; Sequence 34, Application US/08747221B  
; Patent No. 6063610  
; GENERAL INFORMATION:  
; APPLICANT: Silver, Gary W.  
; APPLICANT: Wisniewski, Nancy  
; TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid  
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Carol Talkington Verser, Ph.D.  
; ADDRESSEE: Heska Corporation  
; STREET: 1825 Sharp Point Drive  
; CITY: Fort Collins  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80525  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: WordPerfect for Windows, Version 7.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/747,221B  
; FILING DATE: No. 6063610ember 12, 1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Verser, Carol Talkington  
; REGISTRATION NUMBER: 37,459  
; REFERENCE/DOCKET NUMBER: FC-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 970/493-7272  
; TELEFAX: 970/484-9505  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1785 nucleotides  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1785  
US-08-747-221B-34  
Alignment Scores:  
Pred. No.: 0.209 Length: 1785  
Score: 80.50 Matches: 38  
Percent Similarity: 38.41% Conservative: 25  
Best Local Similarity: 23.17% Mismatches: 62  
Query Match: 11.26% Indels: 39  
Gaps: 8  
US-09-701-618A-4 (1-141) x US-08-747-221B-34 (1-1785)  
Qy 4 GluGluMetCAsnGlyPyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23  
Db 1213 AAAGAACCAAAACGATGACATC-----GAAACCTATGTCCAACTTAAAGGC 1257  
Qy 24 SerLeuGlyAlaValAspAspMetLeuLysThrMetMetAlaValAspArgAsnGluLeu 43

Db 1258 GATGCTGTTATCTCAAGGAATTTACCGTACCTTGAAAGCATATTTTCATGAATC 1317  
Qy 4 LeuGlnLys-----LeuAspProLeuGluGlnAlaLysValAspLeuValSer 59  
Db 1318 AAAAGAAATTCACACTTGTATTTGTATAGGTTATACAGATGATACGTATAGTATATAA 1377  
Qy 60 AlaTyrThrLeu----- 63  
Db 1378 AGTTATATCTTGGCCCTATCCATGAGGGGTTCTTCCAGAGACTTACTCATGGTATGATTTA 1437  
Qy 64 -----AsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsn--- 77  
Db 1438 GGAATCTTTTGGAAACCTTTGGATGTCCTATTTTGGGAACACGACATTTCTATA 1497  
Qy 78 ProLysGluHisProValLysGlnGluLeuArg---IleArgValTyrMetAsnArg 96  
Db 1498 CCGCAGAGAT-----GCTATGACAGACTGTGAAAGATGTCAGAGATCGACCAATTTT 1551  
Qy 97 ValLys-----GluIleThrAspLysLysValAlaLysLeuAspArgGlyAlaAla 114  
Db 1552 GTAAAGATGGAACCTACATCAACACCTGAAGATGATCATGTGAT-----ACA 1602  
Qy 115 SerArgPheValIySlybAlaLeuTrpGluProLysArgIysSerThrProLysValAla 134  
Db 1603 AAAAGACATTTTAAACGACATTTTGGGACCATACACGACGAAGAACCAAAATATTG 1662  
Qy 135 AsnLysGlyLys 138  
Db 1663 GACATGGGAAAA 1674  
RESULT 13  
US-08-747-221B-35/C  
; Sequence 35, Application US/08747221B  
; Patent No. 6063610  
; GENERAL INFORMATION:  
; APPLICANT: Silver, Gary W.  
; APPLICANT: Wisniewski, Nancy  
; TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid  
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Carol Talkington Verser, Ph.D.  
; ADDRESSEE: Heska Corporation  
; STREET: 1825 Sharp Point Drive  
; CITY: Fort Collins  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80525  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: WordPerfect for Windows, Version 7.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/747,221B  
; FILING DATE: No. 6063610ember 12, 1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Verser, Carol Talkington  
; REGISTRATION NUMBER: 37,459  
; REFERENCE/DOCKET NUMBER: FC-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 970/493-7272  
; TELEFAX: 970/484-9505  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1785 nucleotides  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-747-221B-35

## Alignment Scores:

Pred. No.: 0.209 Length: 1785  
Score: 80.50 Matches: 38  
Percent Similarity: 38.41% Conservative: 25  
Best Local Similarity: 23.17% Mismatches: 62  
Query Match: 11.26% Indels: 39  
DB: 3 Gaps: 8

US-09-701-618A-4 (1-141) x US-08-747-221B-35 (1-1785)

```
QY 4 GluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 573 AAAGAACAAAACAGTGCATC-----GAAGCCATGTCCAACTAAAGGC 529
QY 24 SerLeuGlyAlaValAspAspMetLeuLysThrMetMetAlaValSerArgSerGluLeu 43
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 528 GATGCTGTTATCTCCAGAGATTACCGTACCTGAAAGCCATATTTTCAATGAAATC 469
QY 44 LeuGlnLys-----LeuAspProLeuGluGlnAlaLysValAspLeuValSer 59
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 468 AAAAGAAATTCCAACTGTATTTGTATAGTTATCAGATGATACGATAGTATATAA 409
QY 60 AlaTyrThrLeu----- 63
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 408 AGTTATATCTTGCCCATATGATGGGTTCTTGCAGAGATTAGTCATGATGATTTA 349
QY 64 -----AsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsn--- 77
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 348 GGATATCTTTTGGCAAACTCTTGGATGTTCTTATTTGGGAACAACCATTTCTATA 289
QY 78 ProLysGluHisProValLysGlnGluLeuArg---IleArgValTyrMetAsnArg 96
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 288 CCGCAAGAT-----GCTATGCAAGCTCTGGAAGAGATGTCAGATCTGCAACAATTTT 235
QY 97 ValLys-----GluIleThrAspLysLysAlaLysLysLeuAspArgGlyAlaAla 114
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 234 GTAAAGAAATGGAACCTACATCAACAACGTAAAGATGATCATGTGAT-----ACA 184
QY 115 SerArgPheValLysAlaLeuTrpGluProLysArgLysSerThrProLysValAla 134
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 163 AAAAAGACATTAAACGACATTTTGGGAAACCATACACAGACGAAGAACCAAAATATTG 124
QY 135 AsnLysGlyLys 138
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 GACATGCGAATA 112
```

## RESULT 14

US-09-005-051-34  
Sequence 34, Application US/09005051  
Patent No. 6291222

## GENERAL INFORMATION:

APPLICANT: Silver, Gary W.  
APPLICANT: Wisniewski, Nancy  
TITLE OF INVENTION: No. 6291222e1 Carboxylesterase Nucleic Acid  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Carol Talkington Verser, Ph. D.  
STREET: 1825 Sharp Point Drive  
CITY: Fort Collins  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80525

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: WordPerfect for Windows, Version 7.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/005,051

FILING DATE:  
CLASSIFICATION:

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/747,221  
FILING DATE: No. 6291222e1, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Verser, Carol Talkington  
REGISTRATION NUMBER: 37,459  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 970/493-7272  
TELEFAX: 970/484-9505  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1785 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1785

US-09-005-051-34

## Alignment Scores:

Pred. No.: 0.209 Length: 1785  
Score: 80.50 Matches: 38  
Percent Similarity: 38.41% Conservative: 25  
Best Local Similarity: 23.17% Mismatches: 62  
Query Match: 11.26% Indels: 39  
DB: 3 Gaps: 8

US-09-701-618A-4 (1-141) x US-09-005-051-34 (1-1785)

```
QY 4 GluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1213 AAAGAACAAAACAGTGCATC-----GAAGCCATGTCCAACTAAAGGC 1257
QY 24 SerLeuGlyAlaValAspAspMetLeuLysThrMetMetAlaValSerArgSerGluLeu 43
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1258 GATGCTGTTATCTCCAGAGATTACCGTACCTGAAAGCCATATTTTCAATGAAATC 1317
QY 44 LeuGlnLys-----LeuAspProLeuGluGlnAlaLysValAspLeuValSer 59
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1318 AAAAGAAATTCCAACTGTATTTGTATAGTTATCAGATGATGATGATTTA 1377
QY 60 AlaTyrThrLeu----- 63
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1378 AGTTATATCTTGCCCATATGATGGGTTCTTGCAGAGATTAGTCATGATGATTTA 1437
QY 64 -----AsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsn--- 77
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1438 GGATATCTTTTGGCAAACTCTTGGATGTTCTTATTTGGGAACAACCATTTCTATA 1497
QY 78 ProLysGluHisProValLysGlnGluLeuArg---IleArgValTyrMetAsnArg 96
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1498 CCGCAAGAT-----GCTATGCAAGCTCTGGAAGAGATGTCAGATCTGCAACAATTTT 1551
QY 97 ValLys-----GluIleThrAspLysLysAlaLysLysLeuAspArgGlyAlaAla 114
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1552 GTAAAGAAATGGAACCTACATCAACAACGTAAAGATGATCATGTGAT-----ACA 1602
QY 115 SerArgPheValLysAlaLeuTrpGluProLysArgLysSerThrProLysValAla 134
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1603 AAAAGACATTAAACGACATTTTGGGAAACCATACACAGACGAAGAACCAAAATATTG 1662
QY 135 AsnLysGlyLys 138
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1663 GACATGCGAATA 1674
```

## RESULT 15

US-09-005-051-35/c  
Sequence 35, Application US/09005051  
Patent No. 6291222  
GENERAL INFORMATION:



APPLICANT: Silver, Gary W.  
 APPLICANT: Wisniewski, Nancy  
 TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid  
 TITLE OF INVENTION: Molecules, Proteins and Uses Thereof  
 NUMBER OF SEQUENCES: 66  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Carol Talkington Verner, Ph.D.  
 ADDRESSEE: Heeka Corporation  
 STREET: 1825 Sharp Point Drive  
 CITY: Fort Collins  
 STATE: Colorado  
 COUNTRY: USA  
 ZIP: 80525  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: Windows 95  
 SOFTWARE: Wordperfect for Windows, Version 7.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/005, 051  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/747,221  
 FILING DATE: No. 6291222el February 12, 1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Verner, Carol Talkington  
 REGISTRATION NUMBER: 37,459  
 REFERENCE/DOCKET NUMBER: FC-1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 970/484-9505  
 TELEFAX: 970/493-7272  
 INFORMATION FOR SEQ. ID NO. 35:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1785 nucleotides  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 US-09-005-051-35

Alignment Scores:	
Pred. No.:	0.209
Score:	80.50
Percent Similarity:	38.41%
Best Local Similarity:	23.17%
Query Match:	11.26%
DB:	3
	8
Length:	1765
Matches:	38
Conservative:	25
Mismatches:	62
Indels:	39
Gaps:	8

US-09-701-618A-4 (1-141) X US-09-005-051-35 (1-1785)

```

Oy      4  GluGluMetAvnGlnAspTryProValGluIleHisGluSerLeuThrIleValGluSer 23
Db      573  AAlAGAAcAAAGAGATGCATC-----GAGGCTATGTCCAACTTAAAGGC 529p
Oy      24  SerLeuGlnAlaValAspAspMetLeuValThrMetMetAlaValSerArgAsnGluLeu 43
Db      528  GATGCTGGTTATCTCCAGAAATTATACCGTACTTGAAGCCATATTTTTCATGGAATC 465p
Oy      44  LeuGlnIys-----LeuAspProLeuGluGlnAlaIysValAspLeuValSer 59
Db      468  AAAGAAATTCCAACTGTATTGTATAGATTATCATGATGATACGTATAGTATATATAA 409p
Oy      60  AlaTyrThrIleu----- 63
Db      408  AGTTATATCTTGCCTATCGATGAGGGTCTCTTGCCAGAGATTATGATCATGCTGATTTTA 349p
Oy      64  -----AsnSerMetPheIleProValTyrLeuAlaThrGlnGluValAsn--- 77
Db      348  GGAATATCTTTTGGCAACACCTTTGGATGTCTTCATTTTGGGACAAAGCACAATTTGCATA 289p
Oy      78  ProIysGluHisProValIysGlnIleGluGluArg---IleArgValTyrMetAsnArg 96

```

Db CCGCAAGT-----GCTATGCAAGCTCTGGAAGATGTCAGATCTGCACCAATTTT 235  
 288  
 QY 97 ValIleThrAspIySylAlaIySleuhsparGylAlaIa 114  
 234 GTAAAGATGGAACCTTCATCAACACTGAAAGATCATCATGTGAT-----ACA 184  
 QY 115 SerArgPheValIySylAlaLeuTrpGluProIySargIySerThrProIySValAla 134  
 Db 183 AAAAGACATTTTAAACGACATTTTGGGAACCATCAACAGCAGAACCAATATATTTC 124  
 QY 135 AsnIySglIyS 138  
 Db 123 GACATGGAAAA 112

Search completed: January 5, 2005, 16:12:48  
Job time : 76 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 5, 2005, 13:19:23 ; Search time 339.5 Seconds

(without alignments)  
2345.815 Million cell updates/sec

Title: US-09-701-618A-4

Perfect score: 715

Sequence: 1 MAGEEKNEDYPIVEIHSLTA.....LWPKKSKTPKVKANKSKKH 141

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4176236 segs, 2824127955 residues

Total number of hits satisfying chosen parameters: 8352472

Minimum DB seg length: 0  
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DRV=xlp  
-Q=cgnt2\_1/USPTO.spool\_p/US09701618/runat\_05012005\_085840\_24491/app\_query.fasta\_1.654  
-DB=Published Applications NA -QFMT=fastcap -SUFFIX=rmpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pcr -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcr -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOC=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELop=6 -DELEXT=7

Database: Published Applications NA:

1: /cgnt2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgnt2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgnt2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgnt2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgnt2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgnt2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgnt2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgnt2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgnt2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgnt2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgnt2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgnt2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgnt2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgnt2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgnt2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgnt2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgnt2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgnt2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
19: /cgnt2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
20: /cgnt2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
21: /cgnt2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	704	98.5	426	15	US-10-414-692-42
2	646	90.3	1172	10	US-09-873-367C-318
3	634	88.7	586	15	US-10-029-386-1780
4	538.5	75.3	457	16	US-10-242-533A-49006
5	538.5	75.3	457	16	US-10-085-783A-49006
6	235	32.9	204	15	US-10-029-386-15880
7	141	19.7	1302	16	US-10-424-599-94086
8	133	18.6	1099	18	US-10-425-115-137946
9	125.5	17.6	1302	16	US-10-424-599-94086
10	119.5	16.7	1048	17	US-10-425-114-1617
11	119.5	16.7	1048	17	US-10-767-701-10044
12	116	16.2	630	17	US-10-437-963-24791
13	103	14.4	476	13	US-10-027-633-299522
14	103	14.4	476	13	US-10-027-633-299522
15	100.5	14.1	2823	17	US-10-437-963-45565
16	98	13.7	65	10	US-09-908-975-26240
17	94	13.1	227	9	US-09-923-876-2854
18	94	13.1	227	10	US-09-923-876-2854
19	91	12.7	688	16	US-10-424-599-99034
20	89	12.4	60	10	US-09-908-975-7045
21	88.5	12.4	527	17	US-10-767-701-21423
22	87.5	12.2	1685	17	US-10-115-635-53
23	86.5	12.1	1527	16	US-10-767-701-4984
24	86.5	12.1	1527	16	US-10-282-122A-27913
25	84.5	11.8	3750	9	US-09-938-842A-1662
26	84.5	11.8	3750	10	US-09-938-842A-1662
27	84	11.7	1035	10	US-09-906-179A-142
28	84	11.7	1035	10	US-09-906-179A-142
29	84	11.7	1960	17	US-10-437-963-65096
30	84	11.7	13734	16	US-10-287-226-303
31	83.5	11.7	3222	10	US-09-882-227-595
32	83	11.6	236	9	US-09-923-876-2728
33	83	11.6	236	10	US-09-923-876-2728
34	83	11.6	2074	18	US-10-425-115-3512
35	82.5	11.5	7100	9	US-09-932-183A-1
36	82	11.4	414	9	US-09-960-352-6162
37	81.5	11.4	1350	16	US-10-282-122A-22828
38	81.5	11.4	2080	18	US-10-425-115-47599
39	80.5	11.3	1710	16	US-10-678-521-33
40	80.5	11.3	1785	16	US-10-678-521-34
41	80.5	11.3	1785	16	US-10-678-521-35
42	80.5	11.3	2457	17	US-10-437-963-64138
43	80.5	11.3	2801	16	US-10-678-521-30
44	80.5	11.3	2801	16	US-10-678-521-32
45	80.5	11.3	3285	15	US-10-032-585-6554

## ALIGNMENTS

RESULT 1  
US-10-414-692-42  
; Sequence 42, Application US/10414692  
; Publication No. US20030228607A1  
; GENERAL INFORMATION:  
; APPLICANT: X-Ceptor Therapeutics, Inc.  
; TITLE OF INVENTION: Screening method and modulators having an improved therapeutic  
; FILE REFERENCE: 8012-002-US  
; CURRENT APPLICATION NUMBER: US/10/414, 692  
; PRIOR FILING DATE: 2003-04-14  
; PRIOR APPLICATION NUMBER: 60/372, 650  
; NUMBER OF SEQ ID NOS: 86  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 42  
; LENGTH: 426  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-414-692-42

## Alignment Scores:

Pred. No.: 1.78e-79 Length: 426  
Score: 704.00 Matches: 139  
Percent Similarity: 98.58% Conservative: 0  
Best Local Similarity: 98.58% Mismatches: 2  
Query Match: 98.46% Indels: 0  
DB: 15 Gaps: 0

US-09-701-618a-4 (1-141) x US-10-414-692-42 (1-426)

QY 1 MetAAGIYGIUGIUMeAenGluAAsPTy-PrOVAIGIUIIehISGIUSerLeuThra1a 20  
Db 1 ATGGCAGGTGAAGAAATGAATGAAGATTATCCGTAATAATTCACAGCTTTAAACAGCC 60  
QY 21 LeuGIUSerSerLeuGIYAlaValAspAPMeLeuLysThrMetAlaValSerArg 40  
Db 61 CTGAGAGCTCCCTGGGAGCTGTGGATGACATGCTGAMGCCATGATGCTGTTCTAGA 120  
QY 41 AsnGIUleuLeuGIUnlySLeuAAsPProLeuGIUGIAlaValValAspLeuValSerAla 60  
Db 121 AATGAGTTGTTCAGAAAGTTGACCCATTGGACCAAGCAAGAGTGATTTAGTTCTGCA 180  
QY 61 TyrThrLeuAsnSerMetPheTTPValTYrLeuAlaThrGInGlyValAsnProlySGlu 80  
Db 181 TAACTTAATTAATTCATATGTTTGGGTTTATTGGCAACTCAAGAGTTAATCCAAAAGG 240  
QY 81 HisProValIySGInGIUleuGIUAArgIleArgValTYrMetAsnArgValIySGluIle 100  
Db 241 CATCCAGTGAAGCAGAACTGGAAAGATCAAGGTCTACATACACAGATTAAAGAAATA 300  
QY 101 ThrAspLyIySGInGIYAlaValIySLeuAAsPArgIYAlaValIySerArgPheValIySGlu 120  
Db 301 ACAGACAAAGAAAGAGCTGCCAAGCTGCAAGAGGCTGCTTCAGATTGTCAAGAAC 360  
QY 121 AlaLeuTTrGIUProLyArgIySerThrProLyValAlaAsnLySGIlySserLyS 140  
Db 361 GCACCTGGGAACCCAAAGCAAAAGCACCAACCAAAAGTGGTAATTAAGGAGAAAGCAA 420  
QY 141 His 141  
Db 421 CAC 423

## RESULT 2

US-09-873-367C-318  
; Sequence 318, Application US/09873367C  
; Publication No. US20030165839A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; APPLICANT: Soppet, Daniel  
; APPLICANT: Andress, Gregory  
; APPLICANT: Augustus, Meena  
; APPLICANT: Ebner, Reinhard  
; APPLICANT: Carter, Kenneth  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using  
; FILE REFERENCE: 689290-64  
; CURRENT APPLICATION NUMBER: US/09/873,367C  
; PRIOR FILING DATE: 2003-04-29  
; PRIOR APPLICATION NUMBER: U.S. 60/236,891  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: U.S. 60/236,842  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: U.S. 60/244,867  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: U.S. 60/245,084  
; NUMBER OF SEQ ID NOS: 1067  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 318  
; LENGTH: 1172  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-873-367C-318

## Alignment Scores:

Pred. No.: 1.55e-71 Length: 1172  
Score: 646.00 Matches: 126  
Percent Similarity: 95.00% Conservative: 7  
Best Local Similarity: 90.00% Mismatches: 7  
Query Match: 90.35% Indels: 0  
DB: 10 Gaps: 0

US-09-701-618a-4 (1-141) x US-09-873-367C-318 (1-1172)

QY 1 MetAAGIYGIUGIUMeAenGluAAsPTy-PrOVAIGIUIIehISGIUSerLeuThra1a 20  
Db 118 ATGGCAGGTGAAGAAATGAATGAAGATTATCCGTAATAATTCACAGATTGTGACGC 177  
QY 21 LeuGIUSerSerLeuGIYAlaValAspAPMeLeuLysThrMetAlaValSerArg 40  
Db 178 TTTGAGAAATTCATATGCTGCTGTGATGAGATGCTGAMGCCATGATGCTGTTCTAGA 237  
QY 41 AsnGIUleuLeuGIUnlySLeuAAsPProLeuGIUGIAlaValValAspLeuValSerAla 60  
Db 238 AATGAGTTGTTCAGAAAGTTGACCCATTGGACCAAGCAAGAGTGATTTGTTCTGCA 297  
QY 61 TyrThrLeuAsnSerMetPheTTPValTYrLeuAlaThrGInGlyValAsnProlySGlu 80  
Db 298 TACACATTAATTAATTCATATGTTTGGGTTTATTGGCAACCAAGAGTTAATCTTAAGAA 357  
QY 81 HisProValIySGInGIUleuGIUAArgIleArgValTYrMetAsnArgValIySGluIle 100  
Db 358 CATCCAGTGAAGCAGAACTGGAAAGATCAAGGTCTACATACACAGATTAAAGAAATA 417  
QY 101 ThrAspLyIySGInGIYAlaValIySLeuAAsPArgIYAlaValIySerArgPheValIySGlu 120  
Db 418 ACAGACAAAGAAAGAGCTGCCAAGCTGCAAGAGTGCACGCTTCAAGATTGTAAATAAT 477  
QY 121 AlaLeuTTrGIUProLyArgIySerThrProLyValAlaAsnLySGIlySserLyS 140  
Db 478 GCCCTGGGAACCCAAAGCAAAAGTGCATCAAAAGTTGGCAATTAAGGAGAAAGTAA 537

## RESULT 3

US-10-029-386-1780/C  
; Sequence 1780, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Hanzel, David R.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
; FILE REFERENCE: AEMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; PRIOR FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 1780  
; LENGTH: 586  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO CHR10.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1  
; OTHER INFORMATION: NT HIT: g114743475, EVALUE 0.00e+00  
; OTHER INFORMATION: EST HUMAN HIT: B697814.1, EVALUE 0.00e+00  
; OTHER INFORMATION: SWISSPROT HIT: Q09624, EVALUE 5.70e+00  
US-10-029-386-1780

## Alignment Scores:

Pred. No.: 1.95e-70 Length: 586

Score: 634.00 Matches: 124  
Percent Similarity: 94.24% Conservative: 7  
Best Local Similarity: 89.21% Mismatches: 8  
Query Match: 88.67% Indels: 0  
DB: 15 Gaps: 0

US-09-701-618a-4 (1-141) x US-10-029-386-1780 (1-586)

QY 2 ALAGLVLGLLWETAENGUAPPTTProValGluLLeHISGLUSeRLeuThAlaLeu 21  
DB 515 GCGAGTGAAGAAATTAATGAAGATATCCAGTAAATTCAGAGATTGTCAGCGTTT 456  
QY 22 GLUSeRLeuGLUValAlaValAspAspMetLeuThRMeTcAlaValSerAsn 41  
DB 455 GAGAAATTCATGCTGCTGAGATGAGATGCTGAAGACCAAGATGCTGTTTCTAGAAAT 396  
QY 42 GLUSeRLeuGLUValAlaValAspAspMetLeuGLUValAlaValSerAla 61  
DB 395 GAGTTGTCAGAAAGTTGATCCACTTGAACCAAGAAAGATTGTTTCTGCAATAC 336  
QY 62 ThRLeuAnSerMetPheTTPValTYrLeuAlaThRGLNGLYValaenProLYSGluHis 81  
DB 335 ACATTAATTCAGATTGTTGGTTTATTTGGCAACCAAGAGATTAACTCTTAAGAAACAT 276  
QY 82 ProValLYSGNGLUSeRLeuGLUValAlaValTYrMetAsnArgValLYSGluHis 101  
DB 275 CCAGTAAACAGAAATTCAGAAAGATCATATATGAAACAGATCATAGAAATTAACA 216  
QY 102 AspLYSLYValAlaValLeuAspArgGLYAlaValSerArgPheValLYSGluHis 121  
DB 215 GACAAAGAAAGAGCTGCAAGCTGCAAGAGTCAAGCTTCAAGATTGTTTAAAGAAATGCC 156  
QY 122 LeuTTPGLUProLYSArgLYSLeuThRProLYSValAlaValenLYSGLYSLeuHis 140  
DB 155 CTCGGGAAACCAAAACGAAATATGATCAAAAGTTGCCAATTAAGAAAGAAAGTAAA 99

RESULT 4

US-10-242-535A-49006  
Sequence 49006, Application US/10242535A  
Publication No. US20040013663A1  
GENERAL INFORMATION:  
APPLICANT: ChondroGene Inc.  
APPLICANT: Liew, C.C.  
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
FILE REFERENCE: 4231/2005  
CURRENT APPLICATION NUMBER: US/10/242,535A  
CURRENT FILING DATE: 2002-09-12  
PRIOR APPLICATION NUMBER: US 10/085,783  
PRIOR FILING DATE: 2002-02-28  
PRIOR APPLICATION NUMBER: US 60/305,340  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/275,017  
PRIOR FILING DATE: 2001-03-12  
PRIOR APPLICATION NUMBER: US 60/271,955  
PRIOR FILING DATE: 2001-02-28  
NUMBER OF SEQ ID NOS: 58994  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 49006  
LENGTH: 457  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (52)..(52)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (84)..(84)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (235)..(235)  
OTHER INFORMATION: n is a, c, g, or t

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (381)..(381)  
OTHER INFORMATION: n is a, c, g, or t  
US-10-242-535A-49006

Alignment Scores:

Pred. No.: 1,636-58 Length: 457  
Score: 538.50 Matches: 117  
Percent Similarity: 87.14% Conservative: 5  
Best Local Similarity: 83.57% Mismatches: 18  
Query Match: 75.31% Indels: 3  
DB: 16 Gaps: 0

US-09-701-618a-4 (1-141) x US-10-242-535A-49006 (1-457)

QY 1 MetaLGLVLGLLWETAENGUAPPTTProValGluLLeHISGLUSeRLeuThAla 20  
DB 30 ATGCGAGTGAAGAAATTAATGAAGATATCCAGTAAATTCAGAGATTGTCAGCG 89  
QY 21 LeuGLUSeRLeuGLUValAlaValAspAspMetLeuThRMeTcAlaValSerArg 40  
DB 90 TTTGAGAAATTCATGCTGCTGAGATGAGATGCTGAAGACCAATGATGCTGTTTCTAGA 149  
QY 41 AsnGLUSeRLeuGLUValAlaValAspAspMetLeuGLUValAlaValSerAla 60  
DB 150 AATGAGTTGTCAGAAAGTTGATCCACTTGAACCAAGAAAGATTGTTTCTGCA 209  
QY 61 TYrThRLeuAnSerMetPheTTPValTYrLeuAlaThRGLNGLYValaenProLYSGlu 80  
DB 210 TACCATTAATTCAGATTGTTGGTTTATTTGGCAACCAAGAGATTAACTCTTAAGAA 269  
QY 81 HisProValLYSGNGLUSeRLeuGLUValAlaValTYrMetAsnArgValLYSGluHis 100  
DB 270 CATCCAGTAAACAGAAATTCAGAAAGATCATATATGAAACAGATCATAGAAATTAACA 329  
QY 101 ThRAspLYSLYValAlaValLeuAspArgGLYAlaValSerArgPheValLYSGluHis 120  
DB 330 ACAGACAGAAAGAAAGCT-GGCAAGCTGCAAGAGTCAAGCTTCAAGATTGTTTAAAGAAAT 388  
QY 121 AlaLeuTTPGLUProLYSArgLYSLeuThRProLYSValAlaValenLYSGLYSLeuHis 140  
DB 389 GCCTT-TGGGAAACCAAAACGAAATATGATCAAAAGT-GCCATTAAGAAAGAAAGAAA 445

RESULT 5

US-10-085-783A-49006  
Sequence 49006, Application US/10085783A  
Publication No. US20040037841A1  
GENERAL INFORMATION:  
APPLICANT: ChondroGene Inc.  
APPLICANT: Liew, C.C.  
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
FILE REFERENCE: 4231/2002  
CURRENT APPLICATION NUMBER: US/10/085,783A  
CURRENT FILING DATE: 2002-02-28  
PRIOR APPLICATION NUMBER: US 60/305,340  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/275,017  
PRIOR FILING DATE: 2001-03-12  
PRIOR APPLICATION NUMBER: US 60/271,955  
PRIOR FILING DATE: 2001-02-28  
NUMBER OF SEQ ID NOS: 58994  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 49006  
LENGTH: 457  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (52)..(52)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature

```

; LOCATION: (84)...(84)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (235)...(235)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (381)...(381)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-49006

Alignment Scores:
Pred. No.: 1.63e-58 Length: 457
Score: 538.50 Matches: 117
Best Local Similarity: 87.14% Conservative: 5
Best Local Similarity: 83.57% Mismatches: 18
Query Match: 75.31% Indels: 3
DB: 16 Gaps: 0

US-09-701-618A-4 (1-141) x US-10-085-783A-49006 (1-457)
QY 1 MetAlaGlyGluGluMetCAsnGluAspTyrProValGluIleHISgluSerLeuThra1a 20
DB 30 ATGGCAGGTGAAGAAATTAATGNAAGACTATCCAGTAGAATAATTCAGAGTATTGNCAGCG 89
QY 21 LeuGluSerSerLeuGluValAlaValAspAspMetLeuLysThrMetMetAlaValSerArg 40
DB 90 TTTGAGAAATTCATTTGCTGCTGTGTGATGATGCTGGAAGACCAAGATGCTCTGTTCTAGA 149
QY 41 AsnGluLeuLeuGluLysLeuAspProLeuGluGlnAlaLysValAspLeuValSer1a 60
DB 150 AATGAGTGTGCAAGAAATGATGATCCTTGAACAGCAAAAGGATTTGTTTCTGCA 209
QY 61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGluValAsnProLysGlu 80
DB 210 TACCATTAATTCATATGTTTGGATTAATTTGGCAACCCAGAGTTAACTTAAAGGAA 269
QY 81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
DB 270 CATCAGTAATAACAGAAATTCGAAAGATCAGATATATGACAGACAGCAAGGAAATA 329
QY 101 ThrAspLysLysValAlaAlaLysLeuAspArgGluAlaAsnArgPheValLysLys 120
DB 330 ACAAGCAAGAAAGAGCT-GGCACACTGACAGAGGTGACGCTTCAAGATTGNAAAAAAT 388
QY 121 AlaLeuTrpGluProLysArgLysSerThrProLysValAlaAsnLysGluLysSerLys 140
DB 389 GCCCT-TGGGAACCAACCGAAAAATGCATCAAAAGT--GCCAATTAAGGAAAAAGGAAA 445

RESULT 6
US-10-029-386-15480/c
; Sequence 15480, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15480
; LENGTH: 204
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR10.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1

```

```

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: NT HIT: g14743475, EVALUE 1.00e-111
; OTHER INFORMATION: EST HUMAN HIT: AY63046.1, EVALUE 1.00e-108
; OTHER INFORMATION: SWISSPROT HIT: Q9NQ30, EVALUE 4.20e+00
US-10-029-386-15480

Alignment Scores:
Pred. No.: 1.24e-20 Length: 204
Score: 235.00 Matches: 47
Best Local Similarity: 90.57% Conservative: 1
Best Local Similarity: 88.68% Mismatches: 5
Query Match: 32.87% Indels: 0
DB: 15 Gaps: 0

US-09-701-618A-4 (1-141) x US-10-029-386-15480 (1-204)
QY 88 GluArgGluLeuArgValTyrMetAsnArgValLysGluIleThrAspLysLysAla1a 107
DB 203 GAAGAATCAGAGTATATATGAAACAGATCCAGAAATTAACAGCAAGAAAGCTGGC 144
QY 108 LysLeuAspArgGluValAlaAsnArgPheValLysValAlaLeuTrpGluProLysArg 127
DB 143 AAGCTGACAGAGAGTCAAGCTTCAAGATTGTAAATAATGCCCTCTGGGAACCAAAACCG 84
QY 128 LysSerThrProLysValAlaAsnLysGluLysSerLys 140
DB 83 AAAAATGCATCAAAAGTGGCAATTAAGGAAAAAGTAAA 45

RESULT 7
US-10-424-599-94086
; Sequence 94086, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 94086
; LENGTH: 1302
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_55972C.1
US-10-424-599-94086

Alignment Scores:
Pred. No.: 1.24e-07 Length: 1302
Score: 141.00 Matches: 34
Best Local Similarity: 52.90% Conservative: 39
Best Local Similarity: 24.64% Mismatches: 57
Query Match: 19.72% Indels: 8
DB: 16 Gaps: 2

US-09-701-618A-4 (1-141) x US-10-424-599-94086 (1-1302)
QY 11 ProValGluIleHISgluSerLeuThra1aLeuGluSerSerLeuGluValAlaAspAsp 30
DB 314 CCGAAGCGTTATGATGATGATCAACCACTCTTCCAACTTCAACAACCTCGCAAGC 373
QY 31 MetLeuLysThrMetMetAlaValSerArgAsnGluLeuGluLysLysLeuAspProLeu 50
DB 374 CACTTAACGAGTTTGTCTCTCTCCACCTCAACCTCTCTCCCAATCCCTCTTT 433
QY 51 GluGlnAlaLysValAspLeuValSerAlaTyrThrLeuAsnSerMetPheTrpValTyr 70

```



```

; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53513)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1617
; LENGTH: 874
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700161056_FLI
US-10-425-114-1617
```

```

Alignment Scores:
Pred. No.: 3.72e-05 Length: 874
Score: 119.50 Matches: 34
Percent Similarity: 43.31% Conservative: 21
Best Local Similarity: 26.77% Mismatches: 45
Query Match: 16.71% Indels: 27
DB: 16 Gaps: 3
```

US-09-701-618a-4 (1-141) x US-10-425-114-1617 (1-874)

```

QY 37 AlaValSerArgAsnGluLeuGlnLysLeuAspProLeuGluAlaAlaValasp 56
   31 GCGGCGCGCGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 90
QY 57 LeuValSerAlaTyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyVal 76
   91 CTTCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGT 150
QY 77 AsnProLysGluHisProValLysGlnLysLeuGlnLysGlnLysValLysValTyrMetAsnArg 96
   151 GACCTCGACGAGCACCCCATCAGAAAGAGTTGAAAGTTAAGCTTAATCAGAGAG 210
QY 97 ValLysGluLeuThrAspLysLysLysAla-----AlaLysLeuAspArg 111
   211 TTAATTCATTTGAGACTGGACCAACACCACTCCCTCTACTACCTAAATACA 270
QY 112 GlyAlaAlaSerArgPheValLysLysAlaLeu-----122
   271 CAAGCAGCAGCAGAGTTTATTTGACACTCTTCCATCTGACATCTGATCAGAGAG 330
QY 123 -----TrpGlu-----ProLysArgLysSer 129
   331 AGCATGATGAATTAAGTAGAGAGAAAGCGGAGTTGGTCTGGCGCAGAGAGAAAGCT 390
QY 130 ThrProLysValAlaAlaLeuLys 136
   391 GAACCTTCATGTAAGAAAGAG 411
Db
```

## RESULT 11

```

US-10-767-701-10044
; Sequence 10044, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53513)B
; CURRENT APPLICATION NUMBER: US/10/767, 701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 10044
; LENGTH: 1048
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
```

```

; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS50751_1
US-10-767-701-10044
```

```

Alignment Scores:
Pred. No.: 4.78e-05 Length: 1048
Score: 119.50 Matches: 34
Percent Similarity: 52.63% Conservative: 26
Best Local Similarity: 29.82% Mismatches: 43
Query Match: 16.71% Indels: 11
DB: 17 Gaps: 4
```

US-09-701-618a-4 (1-141) x US-10-767-701-10044 (1-1048)

```

QY 16 GluSerLeuThrAlaLeuGlnSerSerLeuGlyAlaValAspAspMetLeuLysThrMet 35
   123 GAGACGCTCGCGCGCTGCGCATCC-----GTAGGGACCATCTTTCCCATCTTA 170
QY 36 MetAlaValSerArg-----AsnGluLeuGlnLysLeuAspProLeuGlnAla 53
   171 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 230
QY 54 LysValAspLeuValSerAlaTyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThr 73
   231 CGGCGCTTCTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 290
QY 74 GlnGlyValAsnProLysGluHisProValLysGlnLysLeuGlnLysValLysValTyr 93
   291 TCGGGAATGACCTTACAGACCCCATCAGAAAGAAATCGAAAGATTAAGCTTAAG 350
QY 94 MetAsnArgValLys-----GluLeuThrAspLys-----LysLysAlaAlaLys 108
   351 CAGGAAAGATTAATCAATTTAGGCGCTGGCGCAAGGACCACTTCGCTACTACTACA 410
QY 109 LeuAspArgGlyAlaAlaSerArgPheValLysLysAlaLeu 122
   411 ATTAATACACAAAGCAGCAGCAAGTTTATTGACACTCACTT 452
Db
```

## RESULT 12

```

US-10-437-963-24791
; Sequence 24791, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boudharov, Andrey A.
; APPLICANT: Bardazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437, 963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 24791
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_29741C.1
US-10-437-963-24791
```

```

Alignment Scores:
Pred. No.: 6.54e-05 Length: 630
Score: 116.00 Matches: 35
Percent Similarity: 52.14% Conservative: 26
Best Local Similarity: 29.91% Mismatches: 44
Query Match: 16.22% Indels: 12
DB: 17 Gaps: 4
```

US-09-701-618a-4 (1-141) x US-10-437-963-24791 (1-630)



```

Qy      18  LeuThrAlaLeuIuSerLeuGluYla-----ValAspAspMetLeuIysThr 34
Db      52  GTGTCCGCCCCCGAGACACCTTCGGCGGCCGCCAGTCCGTGGGAGACCACTCGCCGAG 111
Qy      35  MetMetAlaValSerArgAsn-----GluLeuLeuGlnLysLeuAspProLeuGluGln 52
Db      112 ATGGTCCGGCGCGCGCGGCCGAGACCCCGACGCCATGCCAGACTCCGCGCGTGGCGCCG 177
Qy      53  AlaIvalAspLeuValSerAlaTyrThrLeuAsnSerMetPheTyrValTyrLeuAla 72
Db      172 GCGGCGCGCTTCTTCGCATGGCGAAGGCGCGCGCTCCCTCTTGCAGCGCAGGAGCTC 231
Qy      73  ThrGln-----GlyValAsnProLysGluHisProValLysGlnGluLeuGluArgIle 90
Db      232 TCGCGGGGTTCGGGAGATTGATCCGGATGGAACCCATTCAAGAAAGAGTTTGAGAGGTTA 291
Qy      91  ArgValTyrMetAsnArgValLysGluIleThrAspLysLysLysAla-----106
Db      292 AGCTGTGGGAGGAAAAGTTAAATCATTTGAGAGCTGGGACAAAGACCAATTGGGCCCA 351
Qy      107 ---AlaLysLeuAspArgGluAlaIleAserArgPheValLysLysAlaLeu 122
Db      352 ACCACATCACTGAATACACAGCAGAGGAAAGTTTCAATTGGTCACTACTATTG 402

```

```

RESULT 13
US-10-027-6132-299522
Sequence 299522, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027.632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 299522
LENGTH: 476
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(476)
OTHER INFORMATION: n = A,T,C or G
US-10-027-6132-299522

```

Alignment Scores:	
Pred. No.:	0.00155
Score:	103.00
Percent Similarity:	95.24%
Best Local Similarity:	95.24%
Query Match:	14,41*
DB:	13
US-09-701-618A-4 (1-141) x US-10-027-633-299522 (1-476)	
	Length: 476
	Matches: 20
	Conservative: 0
	Mismatches: 1
	Indels:
	Gaps: 0

Qy 69 ValTyrLeuAlaThrGlnGlyValAsnProIysGluHisProValIysGlnIleLeuGlu 88  
US-09-101-618A-4 (1-141) x US-10-027-632-299522 (1-476)

Db	141	GTTTATTTTGGCAACCCAGAGTTAATCTTAAGAACATCCAGTAAACAAGAGTTTGGTA	200
Qy	89	Arg	89
Db	201	AGA	203

US-10-027-632-299522  
; Sequence 299522, Application US/10027632  
; Publication No. US20030204075A9

```

? APPLICANT: Wang, David G.
? TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
? TITLE OF INVENTION: Polymorphisms in the Human Genome
? FILE REFERENCE: 108827.129
? CURRENT FILING DATE: 2002-04-30
? PRIOR APPLICATION NUMBER: US 60/218,006
? PRIOR FILING DATE: 2000-07-12
? PRIOR APPLICATION NUMBER: US 60/198,676
? PRIOR FILING DATE: 2000-04-20
? PRIOR APPLICATION NUMBER: US 60/193,483
? PRIOR FILING DATE: 2000-03-29
? PRIOR APPLICATION NUMBER: US 60/185,218
? PRIOR FILING DATE: 2000-02-24
? PRIOR APPLICATION NUMBER: US 60/167,363
? PRIOR FILING DATE: 1999-11-23
? PRIOR APPLICATION NUMBER: US 60/156,358
? PRIOR FILING DATE: 1999-09-28
? PRIOR APPLICATION NUMBER: US 60/146,002
? PRIOR FILING DATE: 1999-08-09
? NUMBER OF SEQ ID NOS: 325720
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 299522
? LENGTH: 476
? TYPE: DNA
? ORGANISM: Human
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(476)
? OTHER INFORMATION: n = A,T,C or G
? US-10-027-632-299522

```

Alignment Scores:	
Pred. No.:	0.00195
Score:	103.00
Percent Similarity:	95.24%
Best Local Similarity:	95.24%
Query Match:	14,41*
DB:	15
Length:	476
Matches:	20
Conservative:	0
Mismatches:	1
Indels:	0
Gaps:	0

US-09-701-618A-4 (1-141) X US-10-027-632-299522 (1-476)

QY 65 ValIYrLeuAlaThGInGInValAsnProIySGInHisProValIySGInGluLeuGlu 88  
 |||||  
 DB 141 GTTATTGTCACCAAGAGTTATCCTAAGACATCCAGTAAACAGAGTTGGTA 200

QY	89	Arg	89
Db	201	AGA	203

RESULT 15  
US-10-437-963-45565  
; Sequence 45565, Application US/10437963  
; Publication No. US20040123343A1

Publication No. US20040123343A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovallc, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei

APPLICANT: Boukharov, Andrey A

